NOAA’s Comprehensive Large Array-data Stewardship System (CLASS)

Wednesday -- 26 October 2005

“Pecora 16” - Global Priorities in Land Remote Sensing
Plenary Session II: Data Availability, Access, and Preservation

Richard G. Reynolds, NOAA/NESDIS
Suitland, Maryland
Topics

- NOAA Data Centers & Mission
- CLASS Vision, Goals, and Overview
- CLASS FY05 Successes
- CLASS FY06 Goals and Plans
- Scope of the CLASS Effort
NOAA’s National Data Centers

- NOAA’s National Data Centers are major archive, access, and assessment sites maintaining, processing, and distributing environmental and geospatial data.

- National Climatic Data Center – [WWW.NCDC.NOAA.GOV](http://WWW.NCDC.NOAA.GOV)
  - Asheville, NC

- National Coastal Data Development Center – [WWW.NCDDC.NOAA.GOV](http://WWW.NCDDC.NOAA.GOV)
  - Stennis, MS

- National Geophysical Data Center – [WWW.NGDC.NOAA.GOV](http://WWW.NGDC.NOAA.GOV)
  - Boulder, CO

- National Oceanographic Data Center – [WWW.NODC.NOAA.GOV](http://WWW.NODC.NOAA.GOV)
  - Silver Spring, MD
NOAA’s National Data Centers
(Continued)

- These Centers provide long-term stewardship for most of NOAA’s environmental and geospatial data, and a broad range of user services.

- They serve as both:
  - Centers of Data -- facilities where extensive collections of given environmental parameter(s) are maintained because of individual or institutional research or operational requirements
  - Agency Record Centers -- facilities where data is made accessible to a large user community, as well as being preserved and protected to certain standards
NOAA’s National Data Centers -- Environmental Data Stewards

Scientific Data Stewardship is ownership, knowledge, utilization, and application of the data

CLASS is the Information Technology infrastructure (hardware and software environment, and tools) underpinning SDS

Data Rescue preserves and makes available historical data sets from obsolete media
CLASS Mission Statement

NOAA's National Data Centers and their world-wide clientele of customers look to CLASS as the sole NOAA IT infrastructure project in which “all” NOAA’s current and future environmental data sets will reside. CLASS provides permanent, secure storage, and safe, efficient data discovery and access between the Data Centers and the customers.
CLASS Vision

Eliminate the various “stove-pipe” systems and produce a unified “enterprise” data access system

----

Centralize NOAA’s numerous data systems for environmental data access.

----

Create a single portal.

----

Retain, as much as possible, portions and modules of existing legacy systems.

----

Be cost-effective
WHY a CLASS?

- Fulfill NOAA’s legal requirement to provide for archive and access to its data
- THE source for the vast majority of observational environmental data generated by NOAA.

- Provide critical products to Customers:
  - Public and Private Research & Development efforts
    - Colleges and Universities
  - Federal, State, and Local Climatologists
  - Agriculture Users, Drought Monitors, and Flood Management
  - Accident Investigators & Legal Community
  - Coastal Monitoring, Algae Blooms, and Fishing Management
CLASS Overview

- CLASS is a web-based data archive and distribution system for NOAA’s environmental data.

- CLASS is an evolving system which will support additional “campaigns,” broader user base, new functionality as implementation continues for the next 10 years.

- CLASS is the principal IT system supporting NOAA’s responsibility as environmental data stewards.
  - CLASS concurrently supports both ongoing operations and new requirements implementation.
CLASS Project Plan

- 10 year Plan
  - Road Map for CLASS Program acquisition
  - Budgetary Funding Requirements for all CLASS elements
  - Life Cycle Planning document
CLASS Statistics (Average Last 12 Months)

- **Ingest** – 71 GB/Day … 26 TB/Year
  - 860,000 Data Sets/Year

- **Distribution (On-line & Subscriptions)** –
  - 44 TB/Year …. 3.63 TB/Month
  - 3,170,000 Data Sets/Year …
    - 263,888 Data Sets/Month
CLASS FY05 Accomplishments

- Revised Summary “10-year” CLASS Project Plan and Budget Requirement
  - $20.8M in FY08 -- $30.4M in FY10 .... $270M over 10-years
- Improved CLASS’s IT Security Posture and Achieved Certification and Accreditation (C&A) for the CLASS System
- Achieved SEI/CMMI Certification at Level-2 for the total Development Team
- Continued 24/7 Operations
- Prepared an MOU with the NASA/IV&V Center in Fairmont
  - Installation Plan Approved
CLASS FY05 Accomplishments
(Continued)

- Ordered equipment to achieve Hardware/Software Commonality among all Nodes
- Planned Suitland Node Relocation to Boulder (NGDC)
- Began ingest of GOES Retrospective Data
  - 920 GBytes/day (20X)
- Established Interface with NMMR
- Conducted NPP/NPOESS Campaign SRR and PDR
- Defined an IDPS to CLASS Interface Control Document
- Working with NASA personnel to define initial requirements to archive EOS/MODIS Level-0 data.
- Began update of the CLASS Long-term Architecture
CLASS FY05 Accomplishments

(Continued)

Operational Software Releases

- CLASS Release 3.1
  - Ingest Enhancements to support IJPS NOAA data
- CLASS Release 3.2
  - Support for Metop-1 data with Readiness for IJPS End-to-End test
  - Subscription for GOES data with Separate GVAR data ‘families;’ GOES-N
  - Upgrade to AIX 5.1/5.2 (64-bit processing structures)
  - Cache Management Enhancements
- CLASS Release 3.3
  - Initial Implementation of Ingest Redesign
  - Upgrades to the Help Pages/Static Pages
  - Map server upgrades;
  - CLASS-NMMR Interface
  - Security enhancements; including capability to deliver data encrypted
    - 3.3.1 … UTC Time utilization

System SAN Capacity Upgrade

- Additional disk space at both CLASS operational sites
- Data Direct Networks … 56 Tbytes (expandable to 302 Tbytes)
- Tape to Disk transfer under way for NGDC Move (90Tbytes of 110 Tbytes)
FY05 CLASS “Studies”

- Long-term Systems Architecture
- Communications Architecture
- CLASS Archive Media (LTO vs. IBM)
- Short-term GOES Retrospective Alternatives
FY06 CLASS Goals & Plans

- Complete the CLASS Long-term Architecture Documents
- Achieve Hardware/Software Commonality among all Nodes
- Relocate Suitland Node to Boulder (NGDC)
  - Establish Multi-node capability
- Support METOP-1 Pre-Launch Testing / Initial Operations
- Complete ingest of GOES Retrospective Data
- NPP “Campaign” Software Releases
  - Finalize NPP Data Submission Agreement
- EOS-MODIS “Campaign” Development and Testing
- Data QA/QC “Campaign” begins
FY06 CLASS Goals & Plans (Continued)

- Establish an interface with NeS
- Enhance and Complete interface with NMMR
- Create an HDF5 Format Compatibility
- Metadata “Campaign” development continues
- Geospatial Capability development begins
- Jason/OSTM “Campaign” development begins

----------------

- Operations Continue
FY06 Hardware/Software Plans

- System Storage Capacity Upgrade
  - LTO-2 to LTO-3 Migration

- CLASS Release 3.4  (Scheduled for November 2005)
  - Security Enhancements
  - C&A Requirements
  - New Hardware & C++ Compiler
  - Multi-sight Accommodations

- CLASS Release 4.0  (Scheduled for March 2006)
  - Basic NPP Support
    - Ingest by File Name
    - Definition of Data Groupings
    - Database Schema Changes
    - Basic Search and Delivery
  - Final IJPS/Metop Release
  - CLASS – NeS Interface
  - Complete CLASS – NMMR Interface
  - Complete Implementation of Ingest Redesign
FY06 Hardware/Software Plans (Continued)

- CLASS Release 4.1  (Scheduled for September 2006)
  - NPP Readiness for NCT-#3
    - Ingest HDF5 Data
    - Create Visualization Images
    - Sub-setting of HDF5 files
    - Identify Cross-reference Information
    - Geographic Search Capabilities
    - Process 2-line Element Files
CLASS -- Challenges for the Future

- Determine exactly which data sets to archive
- Establish API’s for machine-to-machine data exchanges
- Define Extent of User Services to be Provided
- Create Data Distribution Format Standardization or Options
- Launch Reprocessing Methodology
Major CLASS Project
“Functional Campaigns”

- “Core CLASS” Baseline System Development, Expansion, & Evolution
  - FY04-FY16 $94M

- Metadata “Campaign”
  - FY04-FY14 $12M

- QA/QC
  - FY06 ….. $2M/year

- Reprocessing “Campaign”
  - FY09-FY16 $35M

- System O&M
  - FY04 ($2M)-FY14($10M) $11M/year thereafter

Budget numbers are shown in this briefing for the purpose of establishing a reference for relative complexity of a requirement, and level of effort and completeness, and do not represent NOAA, Department of Commerce, or The President’s position regarding specific Congressional Budget Requests.
## Major CLASS Project
### “Data Campaigns”

<table>
<thead>
<tr>
<th>Project</th>
<th>FY Period</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metop-1</td>
<td>FY01-FY07</td>
<td>$6.5M</td>
</tr>
<tr>
<td>NPP</td>
<td>FY04-FY11</td>
<td>$15.8M</td>
</tr>
<tr>
<td>EOS-MODIS</td>
<td>FY04-FY10</td>
<td>$16.7M</td>
</tr>
<tr>
<td>NDE</td>
<td>FY06-FY12</td>
<td>$1.5M</td>
</tr>
<tr>
<td>Insitu</td>
<td>FY06-FY14</td>
<td>$8.0M</td>
</tr>
<tr>
<td>EOS Retrospective</td>
<td>FY07-FY13</td>
<td>$8.2M</td>
</tr>
<tr>
<td>GOES-R</td>
<td>FY07-FY14</td>
<td>$41M</td>
</tr>
<tr>
<td>Metop-2</td>
<td>FY08-FY14</td>
<td>$5.0M</td>
</tr>
<tr>
<td>NPOESS-C1</td>
<td>FY08-FY12</td>
<td>$11.1M</td>
</tr>
<tr>
<td>NEXRAD</td>
<td>FY09-FY15</td>
<td>$8.1M</td>
</tr>
</tbody>
</table>
CLASS Budgets

- FY01: $1.995M
- FY02: $3.599M
- FY03: $2.881M

---------

- FY04: $10.5M
- FY05: $14.6M
- FY06: $10.9M

- FY07: $7.4M *
- FY08: $7.4M *
- FY09: $7.4M *
- FY10+: $7.4M/yr *

* FY07 DOC Direction
THANK YOU!