

An Overview of the NOAA National Data Center and CLASS Landscape

+

CLASS Overview

Kenneth S. Casey, NODC

On Behalf of the

CLASS Operations Working Group (COWG)

+

Kern Witcher, CLASS Program Manager

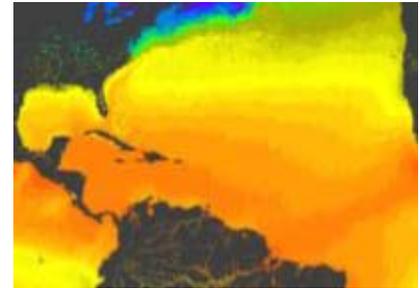
Presentation to the DAARWG, 27 June 2012

First, recall...

The NOAA National Data Centers

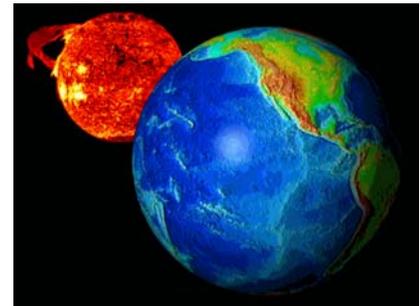
- National Oceanographic Data Center

- Understanding our Oceans and Coasts



- National Geophysical Data Center

- Understanding our World

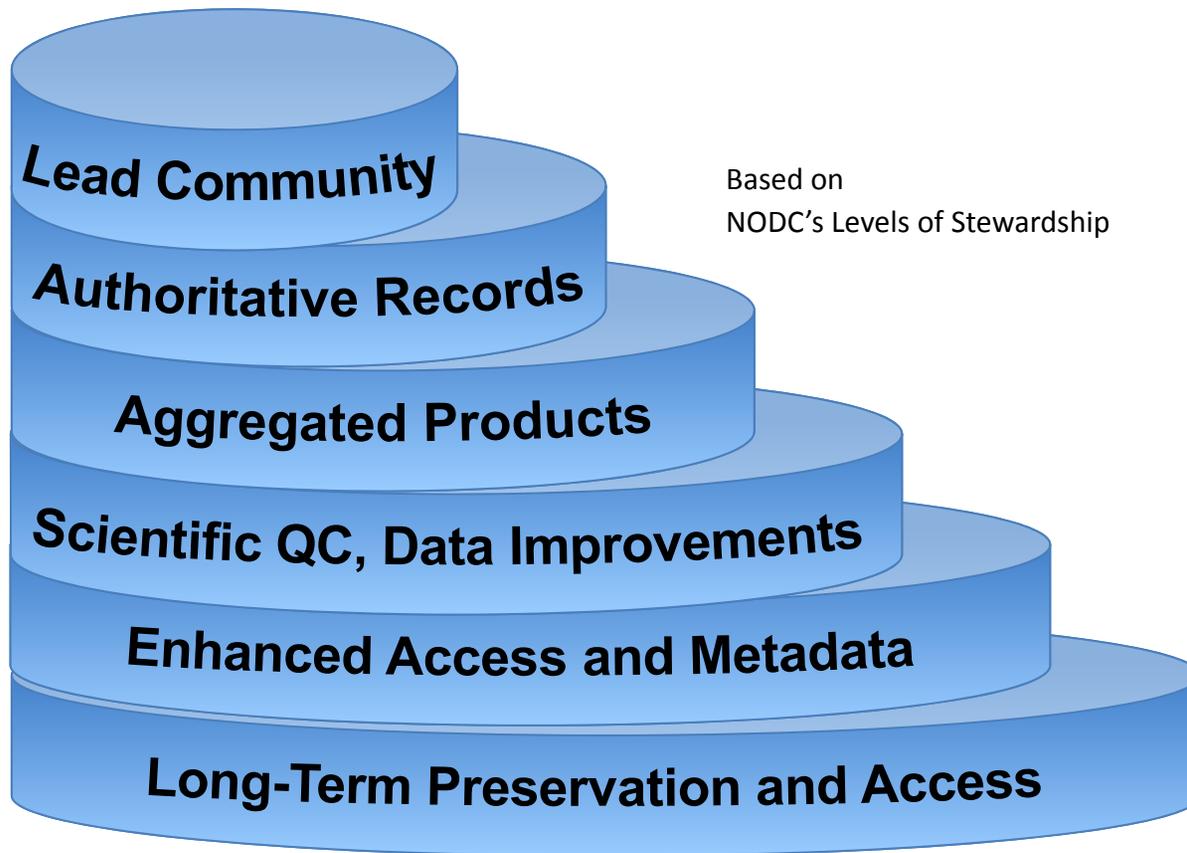


- National Climatic Data Center

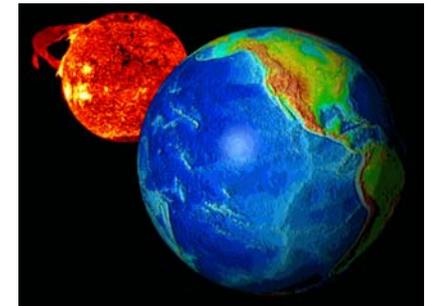
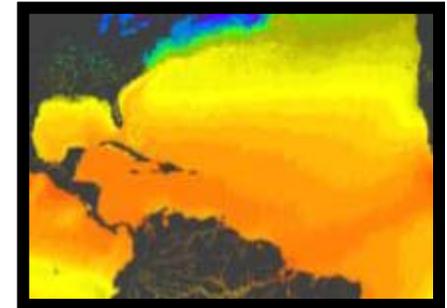
- Understanding our Climate



The NOAA National Data Centers



Based on
NODC's Levels of Stewardship



Across the three Data Centers the words vary a little, but all focus on stewarding environmental data now and for the future.

Comprehensive Large Array-data Stewardship System (CLASS)

- Designed originally for large-volume satellite data sets
- IT infrastructure supporting the lowest level of stewardship
- NESDIS has mandated its use by all three Data Centers



Even the lowest levels of stewardship require non-IT domain knowledge and expertise!

... and the landscape in which we
are operating ...

Consolipalooza...
Consolimaggedon...
Consolitopia...



Consolidation has even hit the popular culture!

Many Forces At Play

Consolidation Free-Body Diagram

- Federal Data Center Consolidation Initiative

FDCCI

- Initiated by the Administration/OMB in 2010
- Reduce hardware, software, real estate, cooling costs

- CLASS Integration into Data Center Operations



- Initiated by NESDIS in 2011
- Support common archival storage needs of the Data Centers

- NESDIS Data Center consolidation



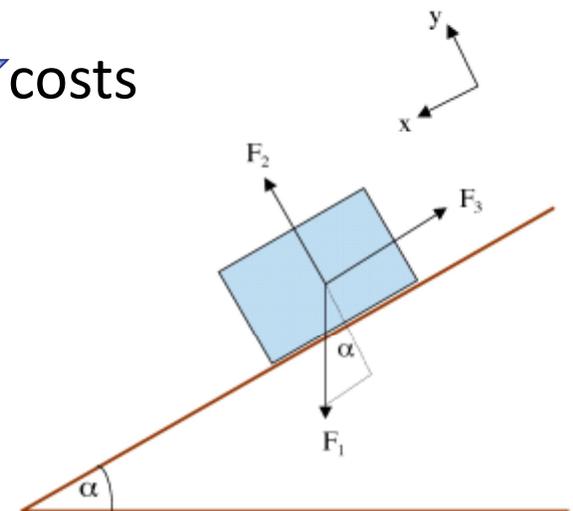
- Initiated by NESDIS in 2012
- Explore options to reduce admin and IT costs

- Other Initiatives

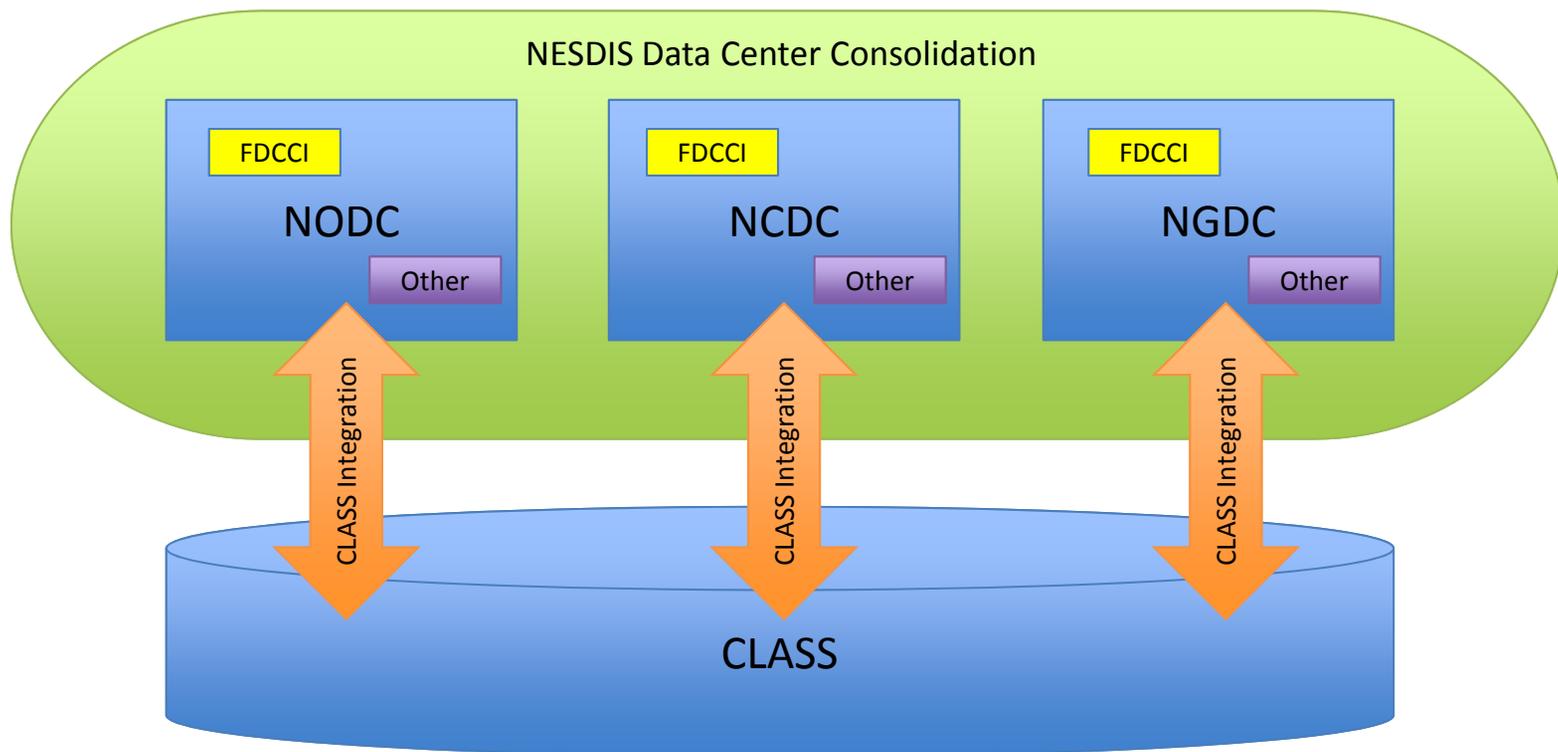
\$

Other

- Initiated by individual Data Centers
- Respond to various pressures



Consolidation Landscape



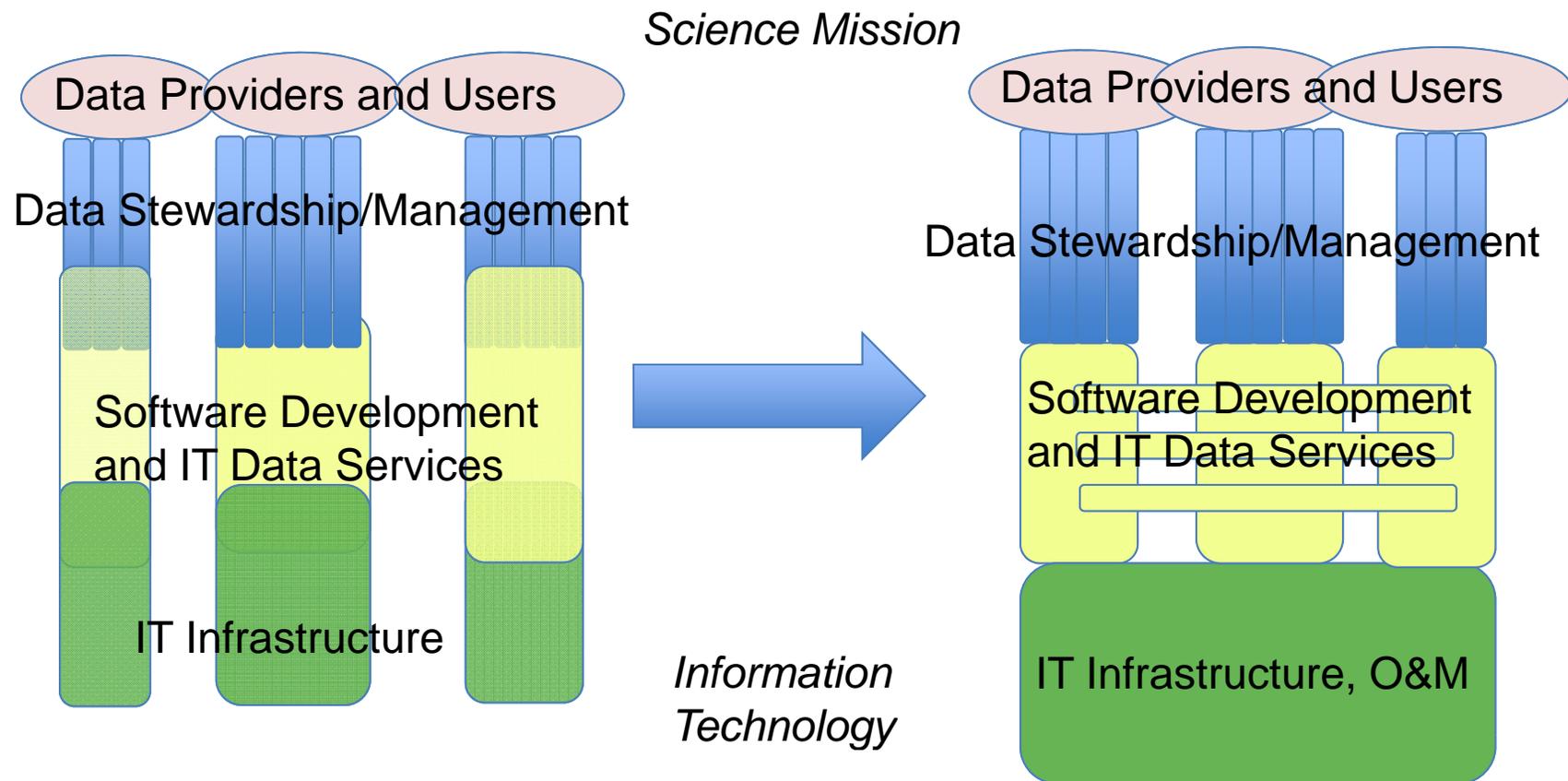
So, what are we doing about it?

Activities Underway...

- FDCCI
 - Server virtualization
 - Reducing IT footprints
 - Data calls/surveys
- NESDIS Data Center Consolidation
 - Deputies meeting to form options
 - Will report to NESDIS Management
- CLASS Integration...

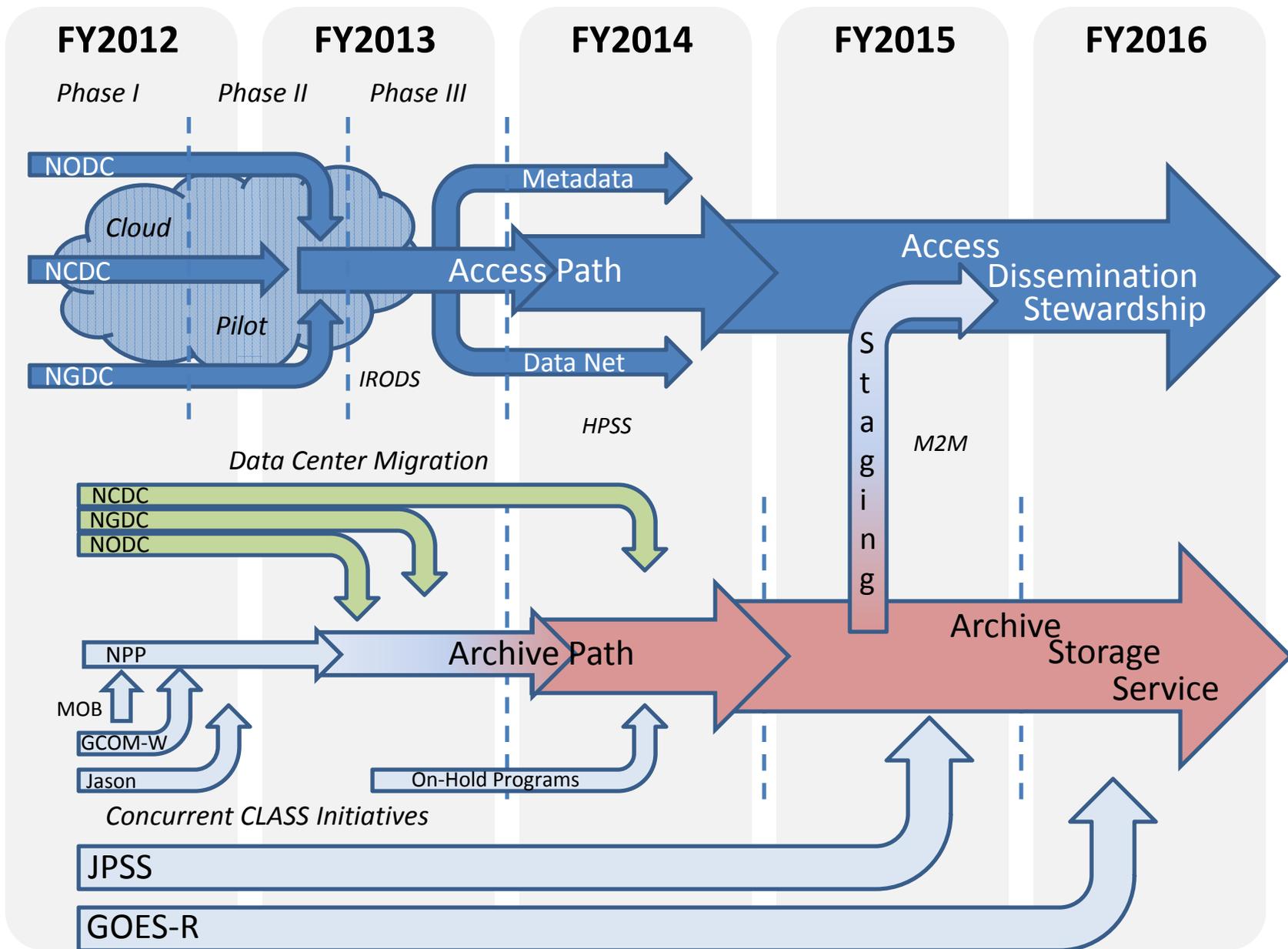
Aerospace Corp. Review

Notional look at the way forward: Containing Costs, Enhancing Mission



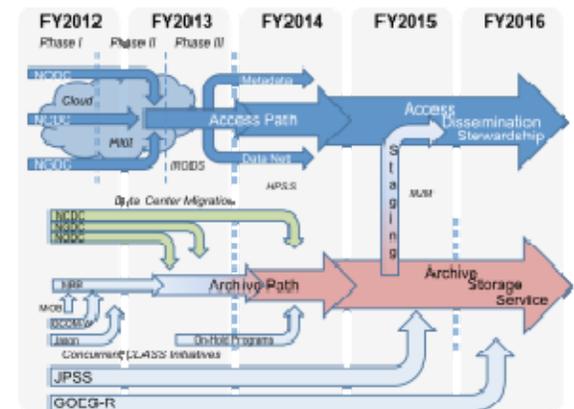
Evolution of the NOAA Archive Architecture

DRAFT – Details under discussion



Data Centers' Data Migration Plan

- Approaches and milestones for integrating CLASS into NOAA Data Center operations by FY15
- Three Phases:
 - **Archival Storage:** use CLASS for safe, long-term storage
 - **Access Services:** expands CLASS to include access capabilities expected by Consumers and functions needed for Data Center stewardship
 - **Operations:** compare levels of service and decommission local Data Center services when appropriate



A metaphor, if you will...



<http://www.dreamstime.com/royalty-free-stock-photo-muddy-boots-image14440895>



<http://www.dreamstime.com/royalty-free-stock-photo-muddy-boots-image14440895>

- Working to integrate CLASS into our archive operations is a bit like our “day job”. You get up, pull on your boots, and make the best of it you can.
- But you are not really very happy in your day job, so you start exploring alternatives. Maybe you take some online classes at night, learn some new skills, invest in a startup... you do something to “change the game”, “live the dream”, or “expand your horizons”... the CLASS Cloud Access Pilot is just that sort of thing.

CLASS Cloud Access Pilot

Why: To test the cost-effectiveness, scalability, performance, and agility of a Cloud solution for access to archived data

Who: Three NOAA Data Centers and CLASS, reported on by CLASS Operations Working Group

What: At least one data collection from each Data Center, including most/all of NODC's data

CLASS Cloud Access Pilot

When: This FY, some overlap into next

Where: A commercial provider of Cloud IaaS: Amazon Web Services (S3/EC2). Government Cloud also being examined.

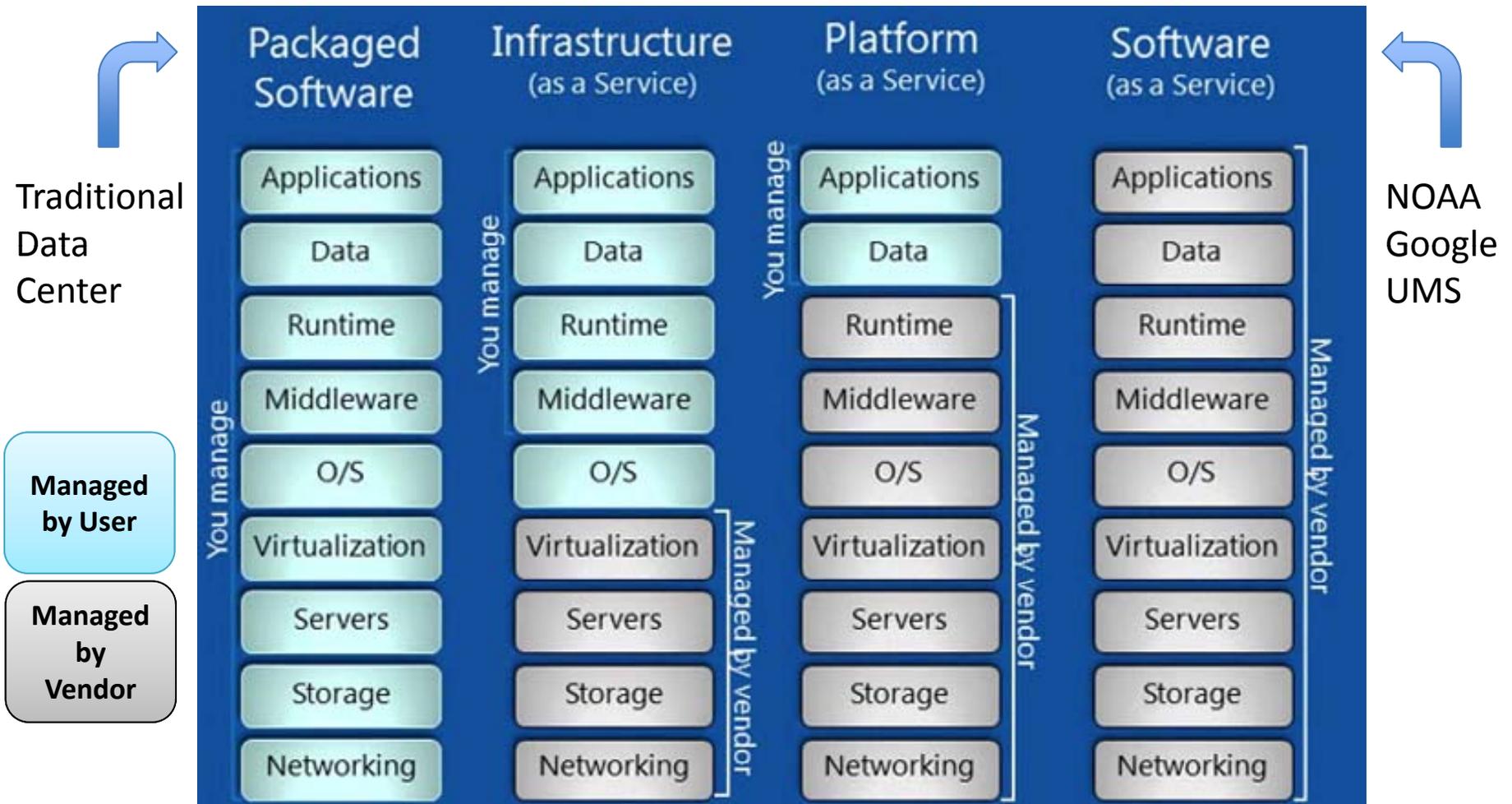
How:

- Three parallel activities
 - Populate Cloud storage with DC-held data
 - Load Data Center access services to the Cloud
 - Populate Cloud storage with CLASS-held data
- Test the Cloud-based data access services with a select group of real-world users

CLASS Cloud Access Pilot

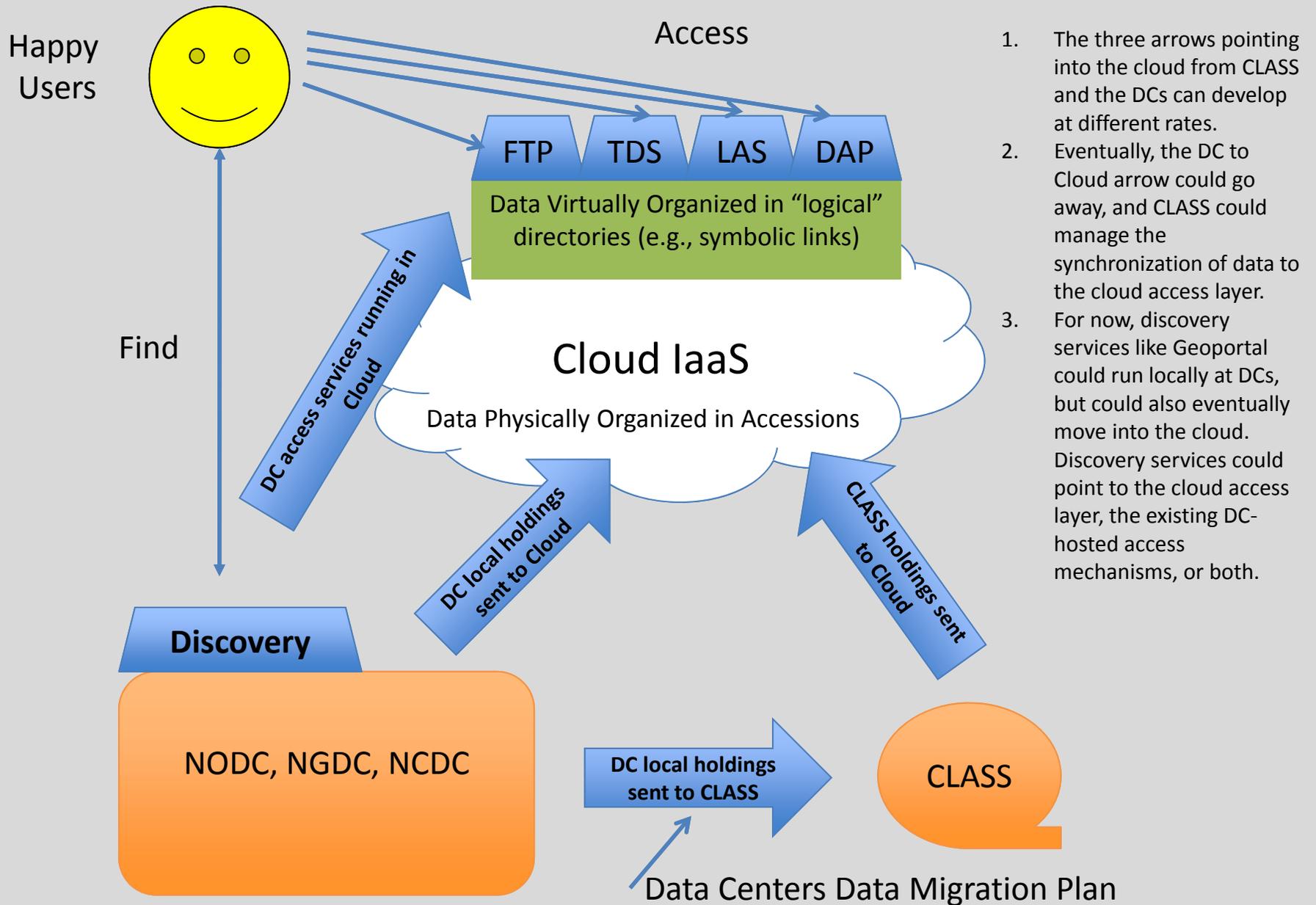
“Common Storage Services using Cloud Technologies”

CLASS Cloud Access Pilot



<http://venturebeat.com/2011/11/14/cloud-iaas-paas-saas/>

CLASS Cloud Access Pilot



1. The three arrows pointing into the cloud from CLASS and the DCs can develop at different rates.
2. Eventually, the DC to Cloud arrow could go away, and CLASS could manage the synchronization of data to the cloud access layer.
3. For now, discovery services like Geoportal could run locally at DCs, but could also eventually move into the cloud. Discovery services could point to the cloud access layer, the existing DC-hosted access mechanisms, or both.

Status in Brief

- Complicated landscape with lots of forces at play
- Progress on *Archival Storage* Phase
- Next few months of CLASS Cloud Access Pilot critical to refine *Services* Phase

Brief Highlights from each NOAA National Data Center

NODC

- Facing 26% proposed FY13 budget cut; examining centralizing IT in MS, admin in MD
- Good progress on Archival Storage phase... on track to have our data in CLASS by end of calendar year
- Excited about CLASS Cloud Access Pilot and running a cloud *computing* pilot as well

NGDC

- Working with CLASS on generic Ingest strategies for efficient data migration
- Working with CLASS on Machine-to-Machine Search & Access Implementations: Data Center interfaces need to talk to CLASS for querying and placing orders
- Working with CLASS on defining data stewardship needs related to the archive holdings

NCDC

- Internal archival storage migration/consolidation plan under way
- Taking the opportunity to “clean up” the 700+ individual datasets in NCDC’s legacy archive
- NCDC to take lead for GOES-R Access function, to lead Access PDR in Spring 2013
- New NCDC website/portals due Summer 2012
- Implementing configuration management of CDRs and other NCDC data products

Now... CLASS Overview...



Comprehensive Large Array-data Stewardship System Overview

Presented to DAARWG

Kern Witcher
CLASS Program Manager
June 27, 2012





Background Refresher





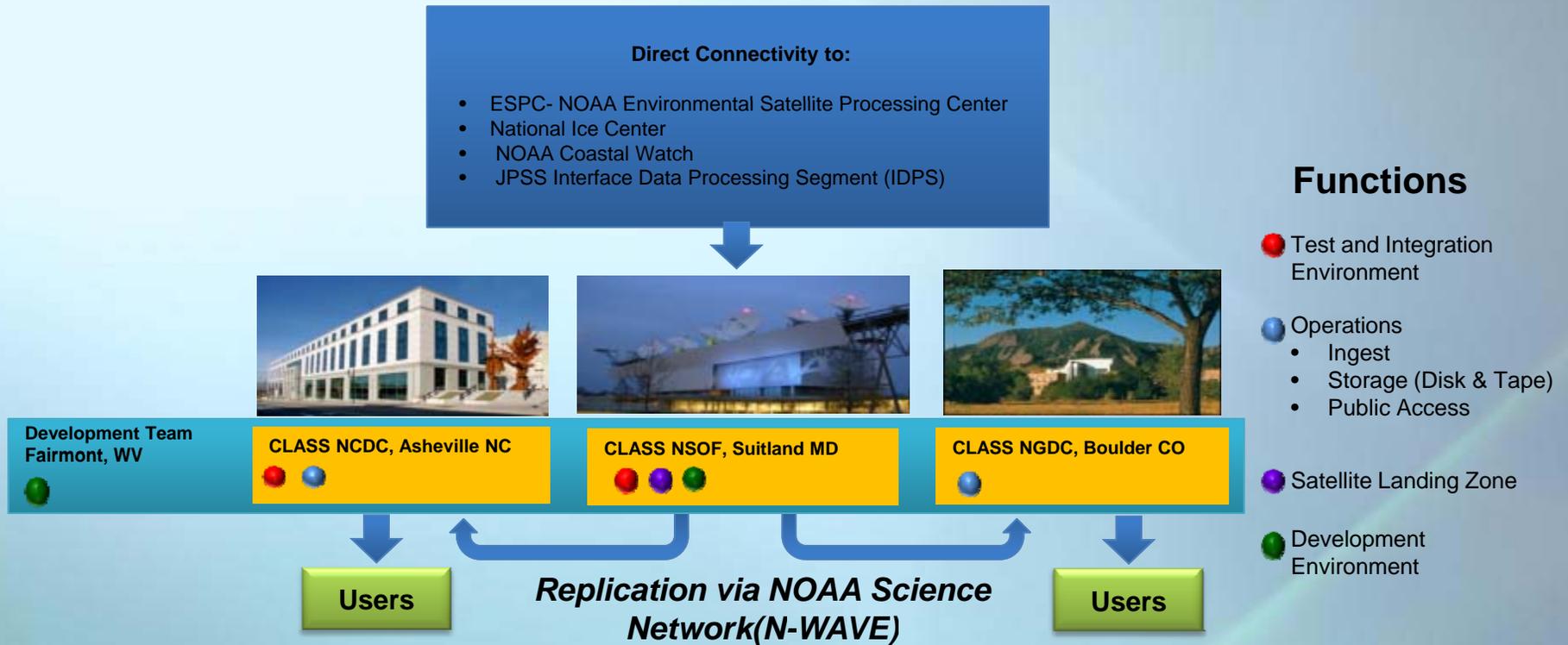
CLASS Level I Requirements (preliminary)

- As an enterprise solution, CLASS will reduce anticipated cost growth associated with storing environmental datasets by:
 - Providing common services for acquisition, security, and project management for the IT system supporting NOAA Archives
 - Consolidating stove-pipe, legacy archival storage (See Note 3 below) systems thereby reducing the number of archival storage related IT projects for NOAA to manage and data systems for customers to access.
 - Relieving data owners of archival storage-related system development and operations issues
- Note 3: Archival storage provides the services and functions for the storage, maintenance and retrieval of archival information packets. Archival storage functions include receiving archival information packets from ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media in which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing archival information packets to access to fulfill orders





Current CLASS Architecture





Recent Accomplishments and Current Efforts





Accomplishments Since Last Briefing

2.0 Core

- Release 6.0 Linux Migration in final Test and Integration. Release is scheduled for July 8,2012
- Completed Server Virtualization Study

3.0 Data Center Migration

- Aerospace completed Phase I ("as is" analysis) of Data Center Con Ops
- Completed Data Center Requirements Documents
- Completed Data Center Interface Control Documents

4.0 JPSS

- Successfully supported the Launch of NPP
- Have ingested 803Tb of NPP data (1.606Pb across both nodes)

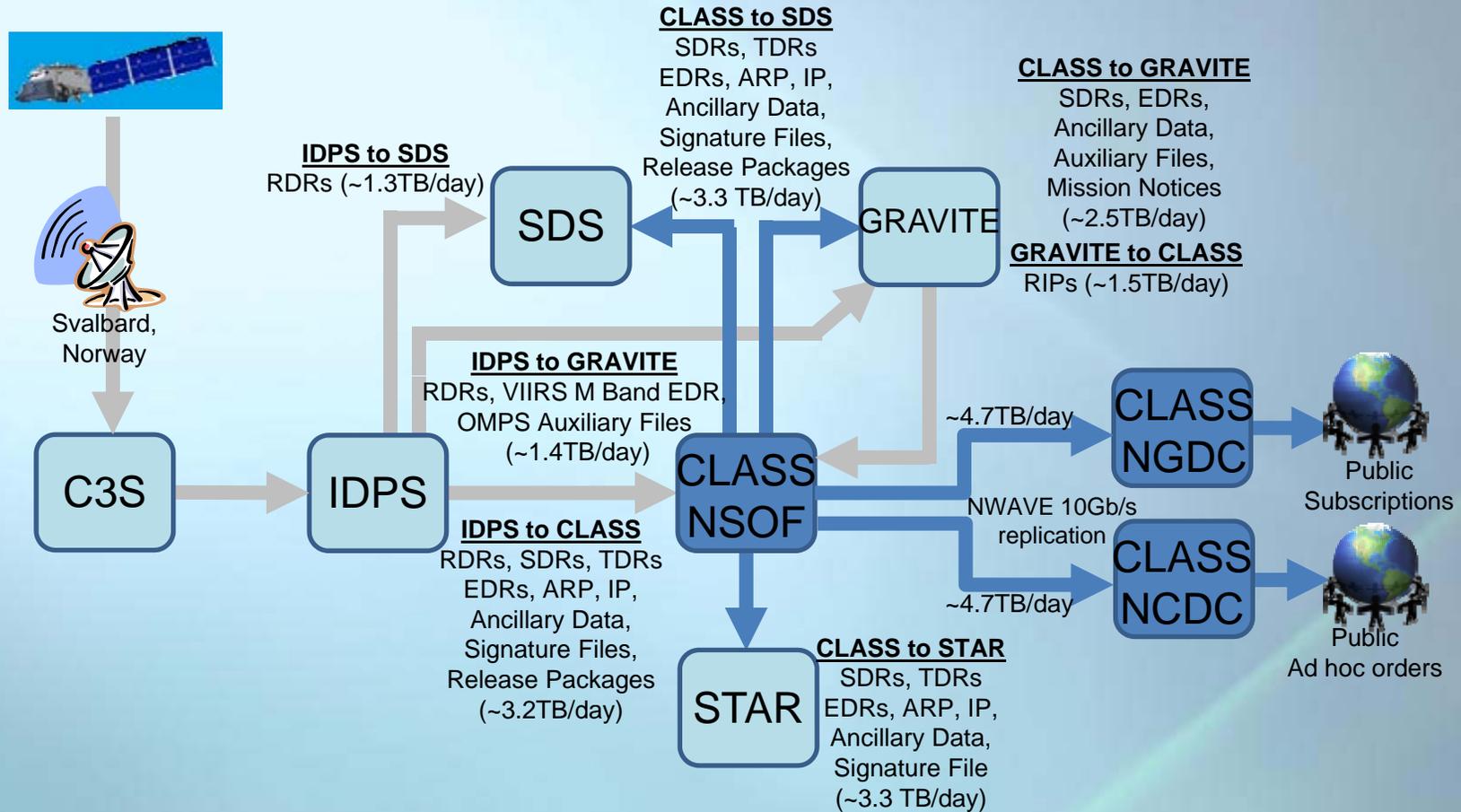
5.0 Goes-R

- Completed Archive and Access PDR
- Completed Interface CDR
- Completed Receipt Node Design
- Machine to Machine (M2M) API in final design
- In Final Stages of HPSS design





CLASS Suomi-NPP Data Flow



NPP Sensors

ATMS – Advance Technology Microwave Sounder
CrIS – Cross-track Infrared Sounder
OMPS – Ozone Mapping and Profiler Suite
VIIRS – Visible and Infrared Imaging Radiometer Suite

NPP Segments

C3S – Command, Control & Communications Segment
GRAVITE – Government Resource for Algorithm Verification, Independent Testing & Evaluation
IDPS - Interface Data Processing Segment
SDS – Science Data Segment
STAR – NOAA Center for Satellite Applications & Research

ARP – Application Related Product
EDR – Environmental Data Record
IP – Intermediate Product
RDR – Raw Data Record
RIP – Retained Intermediate Product
SDR – Sensor Data Record
TDR – Temperature Data Record



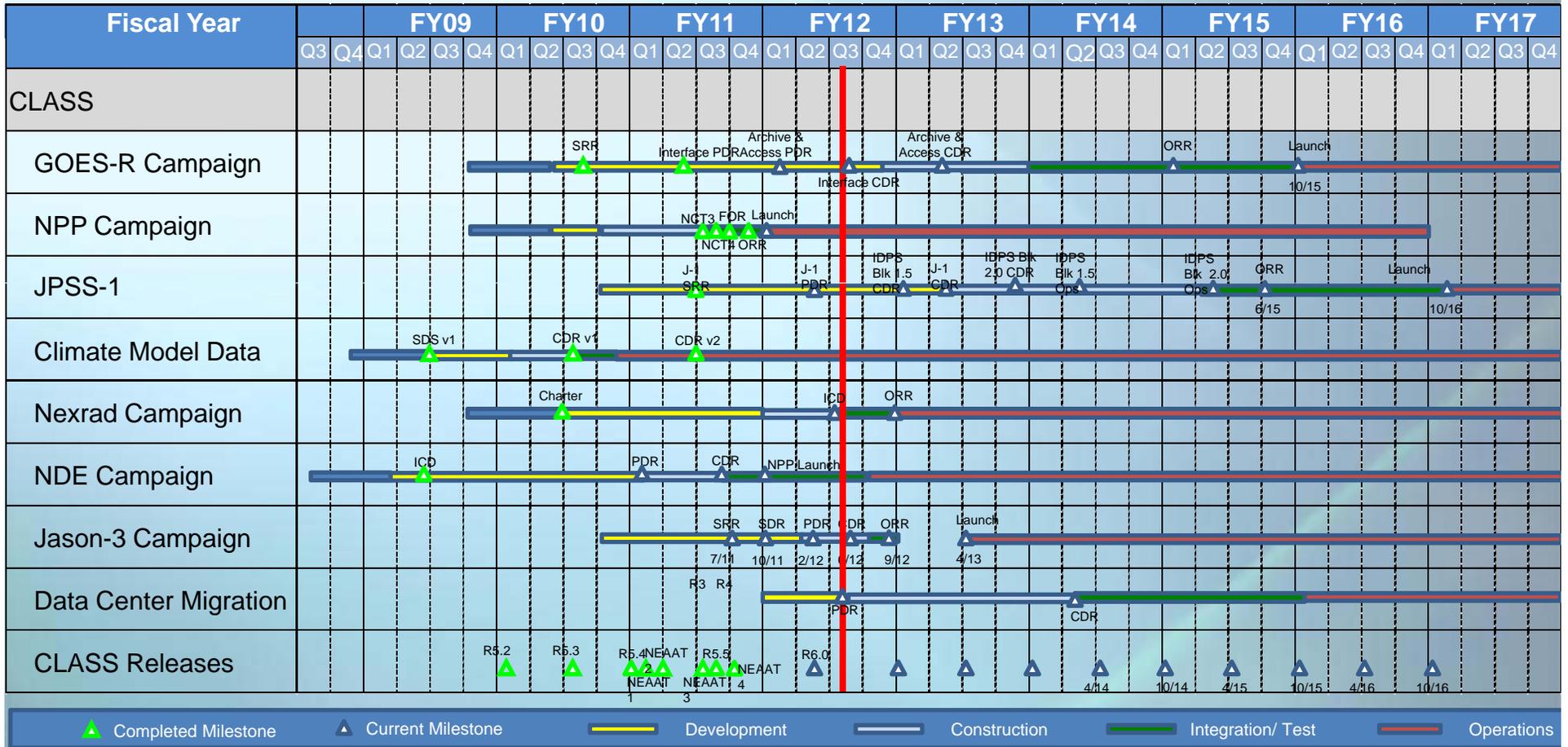


Current Program Projections





CLASS Program Milestones

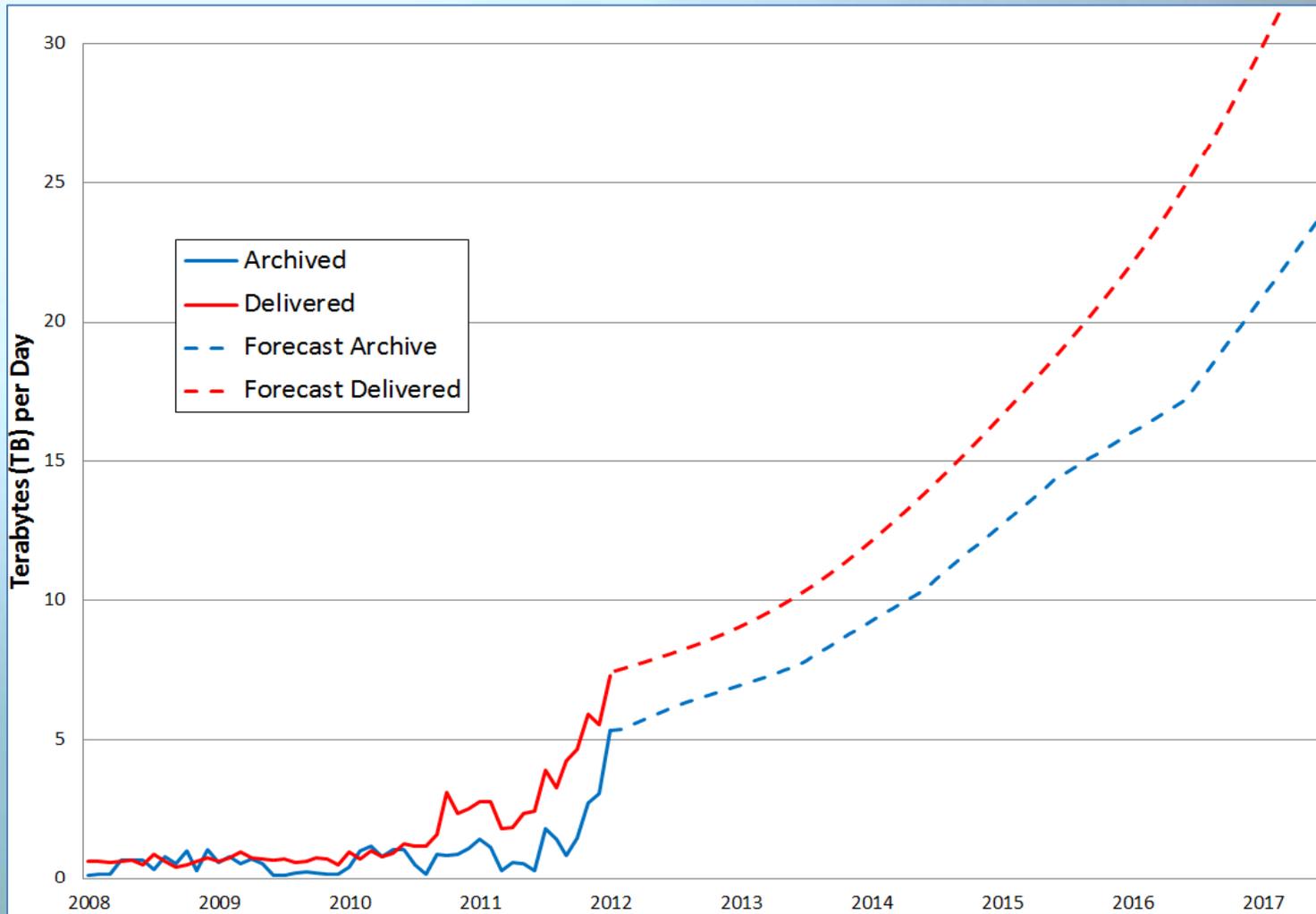


Notes:
 CDR: Critical Design Review
 FOR: Flight Operations Review
 PDR: Preliminary Design Review
 SDR: Systems Definition Review
 SRR: System Requirements Review
 ORR: Operational Readiness Review
 NCT: NPP Connectivity Test
 ICD: Interface Control Document



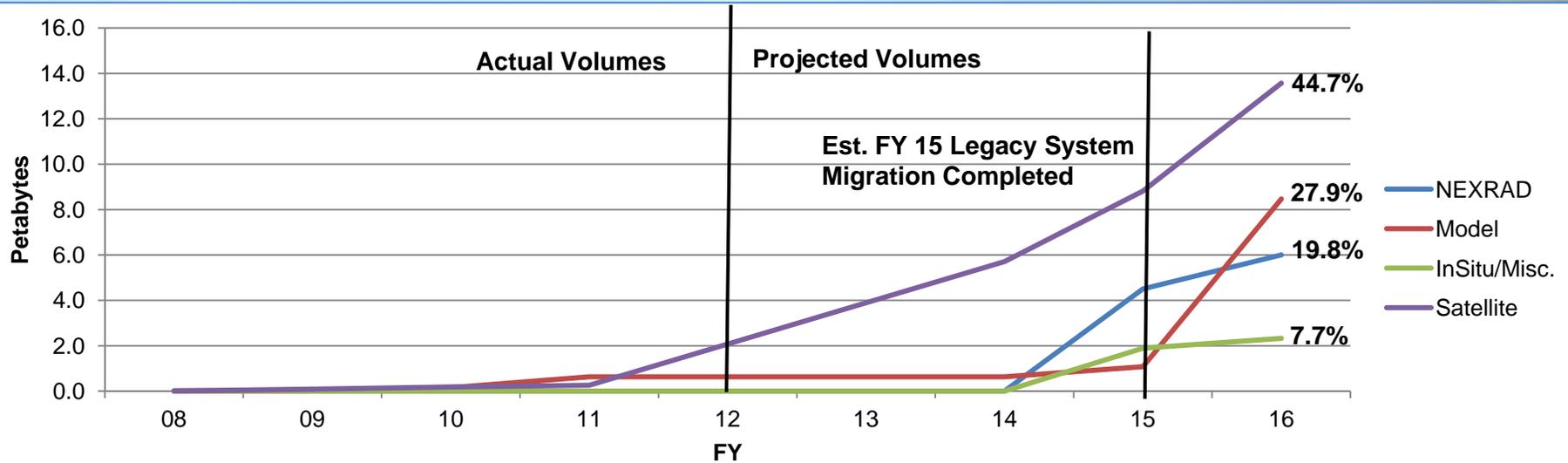
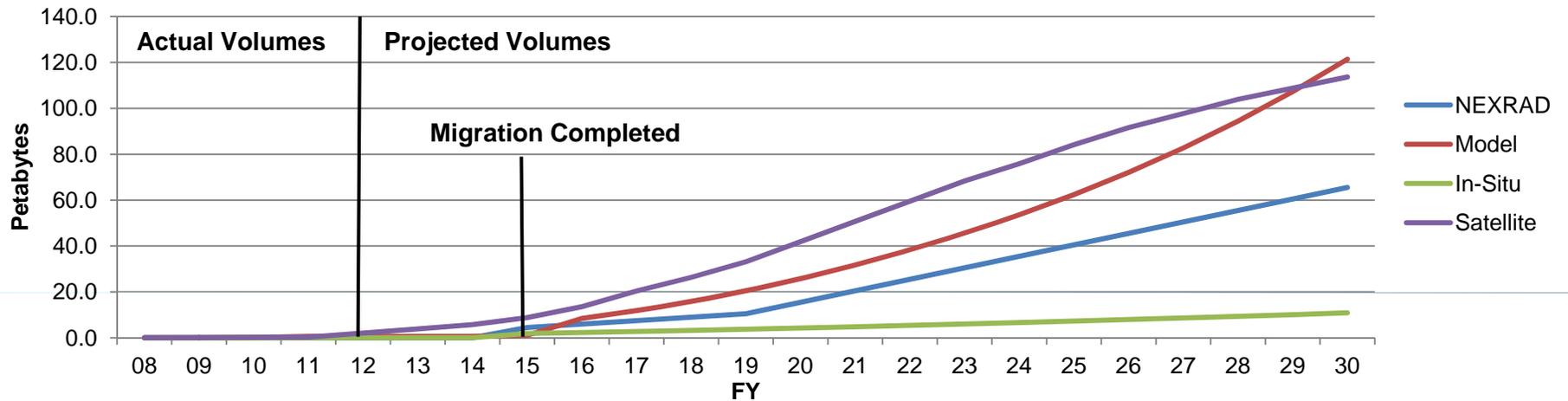


CLASS Projected Ingest and Delivery





NOAA Enterprise Archive System (CLASS) Cumulative Total Volume by Data Type





Near and Long Term Plans





Near-Term Plans

2.0 Core

- Initiate Server Virtualization – Expect at least a 30% reduction of servers.

3.0 Data Center Migration

- Initiate Aerospace Phase II (“to be” state) of Data Center Con Ops
- Complete Cloud Pilot Project
- Complete NexRad Pilot Project

4.0 JPSS

- Update the CLASS IRD and ICD to support the Systems Requirements Review for the IDPS Block 1.5 and 2.0 releases

5.0 Goes-R

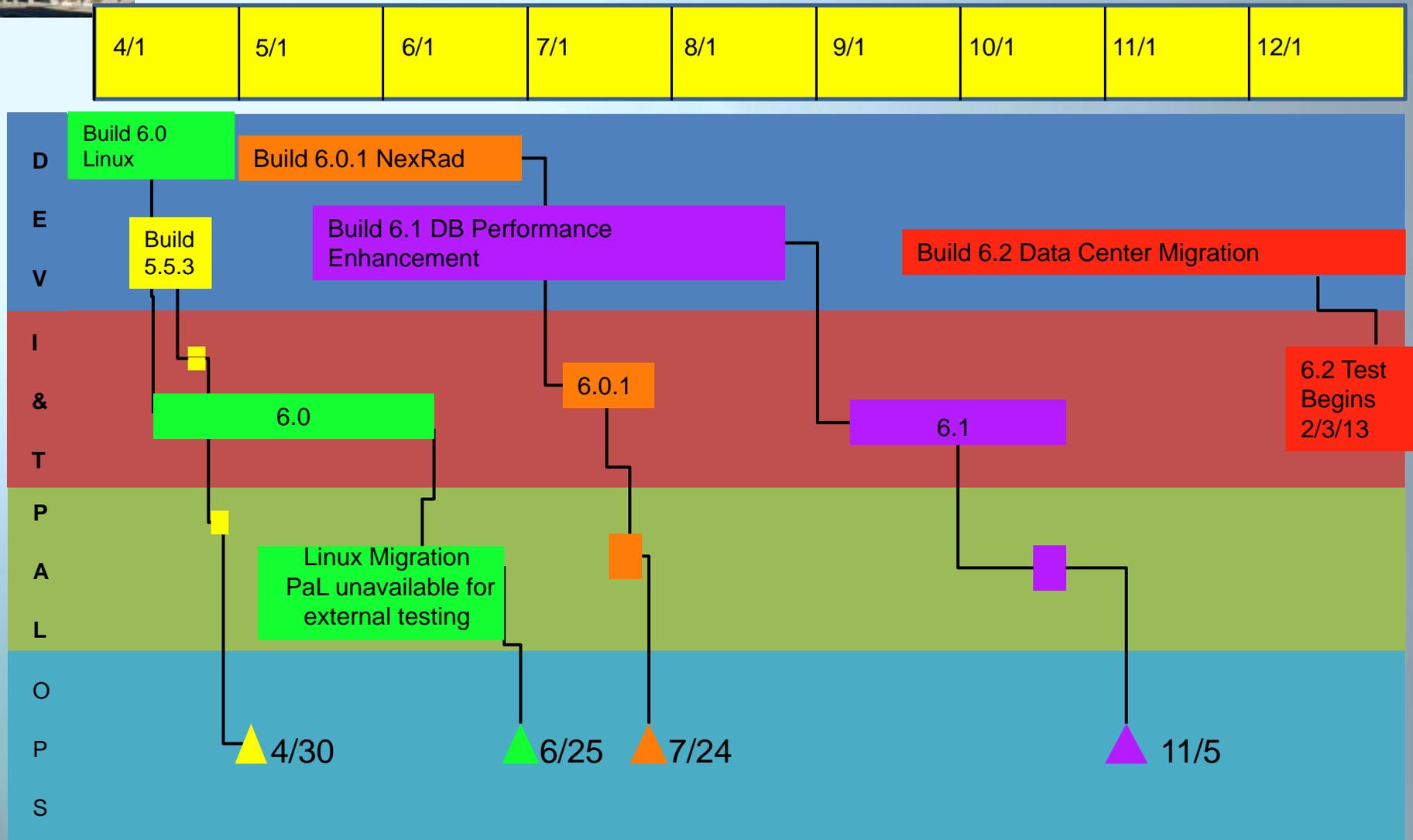
- Complete Archive and Access CDR
- Procure and deploy the GOES-R receipt node into the test and integration environment
- Update HPSS documentation and begin hardware procurement based on IBM recommendations.

- Finalize the M2M design and begin implementation





CY12 Software Releases



5.5.3. NDE, GCOM-W and various NPP fixes

6.0.1 NexRad Pilot

6.2 Data Center Migration

6.0 Linux Transition

6.1 DB Performance enhancements:
Changes to Schema, Updates to M2M

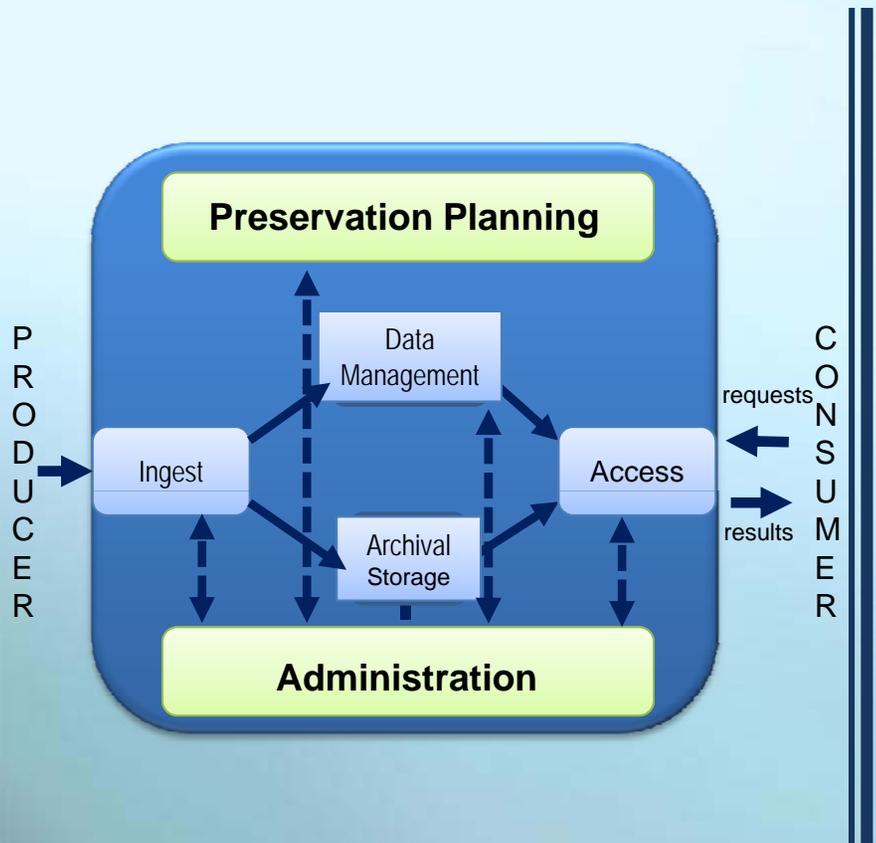
43 Note: Please refer to REMEDY for complete release details



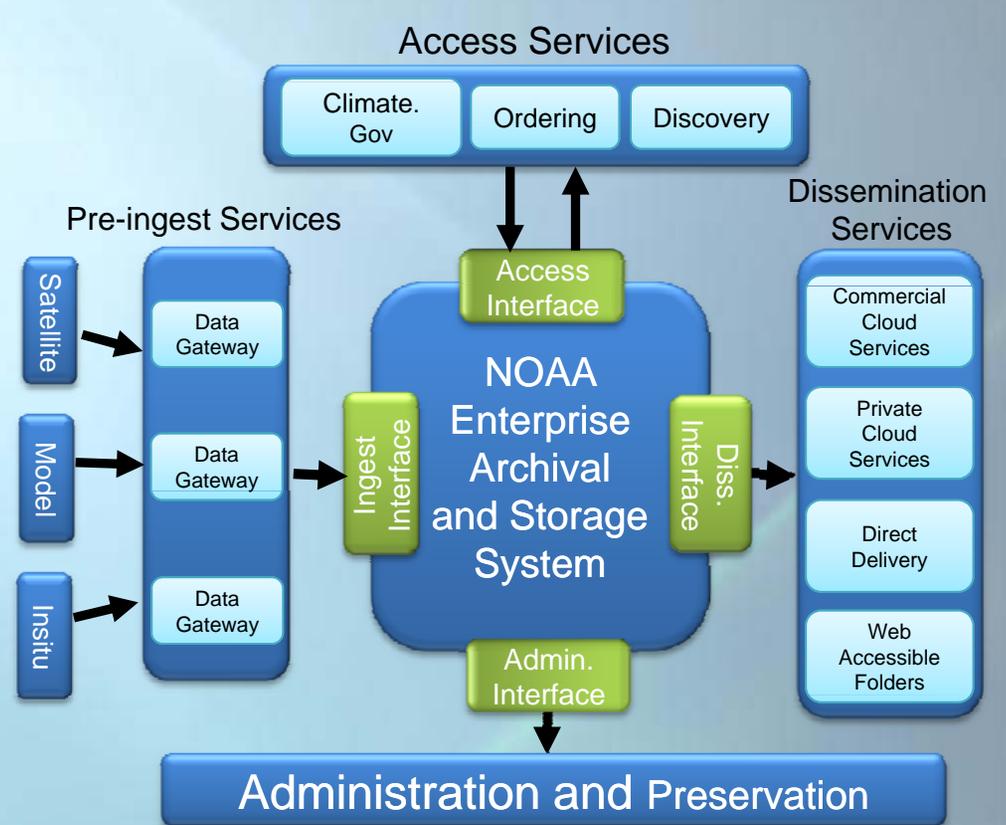


CLASS- Looking Forward

Present CLASS



Future CLASS





Final Thoughts

- CLASS does an outstanding job of meeting it's original intended purpose (Large Satellite Data)
- The program recognizes that for CLASS to become a true NOAA enterprise solution, it must evolve.
- The challenge is how to evolve in a manner that will not disrupt our current operational missions and accomplish this under a constrained budget environment
- DAARWG could be a valuable resource for providing an IV&V function for data and algorithm integrity

