A Review of Improvements and Outstanding Challenges in Data Management and Stewardship by the Center for Satellite Applications and Research (STAR) in support of the NPOESS Preparatory Program (NPP) Satellite Launch

Kathryn A. Shontz
STAR-NDE Liaison
IMSG at NOAA/NESDIS/STAR
Camp Springs, MD 20746

Ingrid Guch, Aleksandar Jelenak and Kent Hughes
NOAA/NESDIS/STAR

Yong Sung Kim
SMRC at NOAA/NESDIS/STAR

Eighth Annual Symposium on Future Operational Environmental Satellites and
Second Conference on Transition of Research to Operations: *Successes, Plans and Challenges*

92nd American Meteorological Society Annual Meeting

January 25, 2012

Outline and Purpose

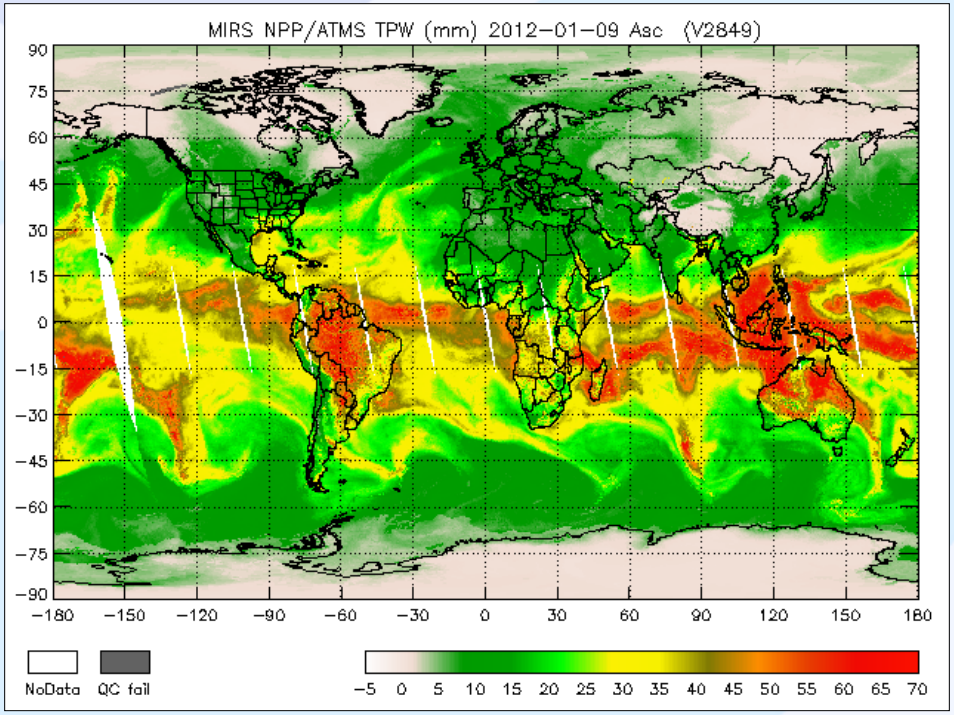
- Background
  - STAR usage of NPP Data
  - Data Management
  - Data Stewardship
- A Year to get NPP Data
  - Scoping Issues
  - Roadmap to Solution
  - Continuing Challenges
- Summary of Work

Goal: *To support the Center for Satellite Applications and Research (STAR) need to get NPOESS Preparatory Project (NPP) data as soon as it becomes available to support NOAA calibration, validation and science monitoring activities in house.*

NPP Data needs at STAR

- NOAA depends on the Center for Satellite Applications and Research (STAR) to be a key part of the calibration and validation of all pre-operational satellites.
- STAR is also tasked to maintain all science properties of NPP after it is transitioned to NOAA from the JPSS Program Office.
- From a long-term perspective, STAR is responsible for creating NOAA Unique Products (NUPs) that are keys to Numerical Weather Prediction and Oceanic and Surface monitoring for the Center.

Thus, STAR will need rapid, consistent and sustained access to NPP data throughout the mission lifetime.



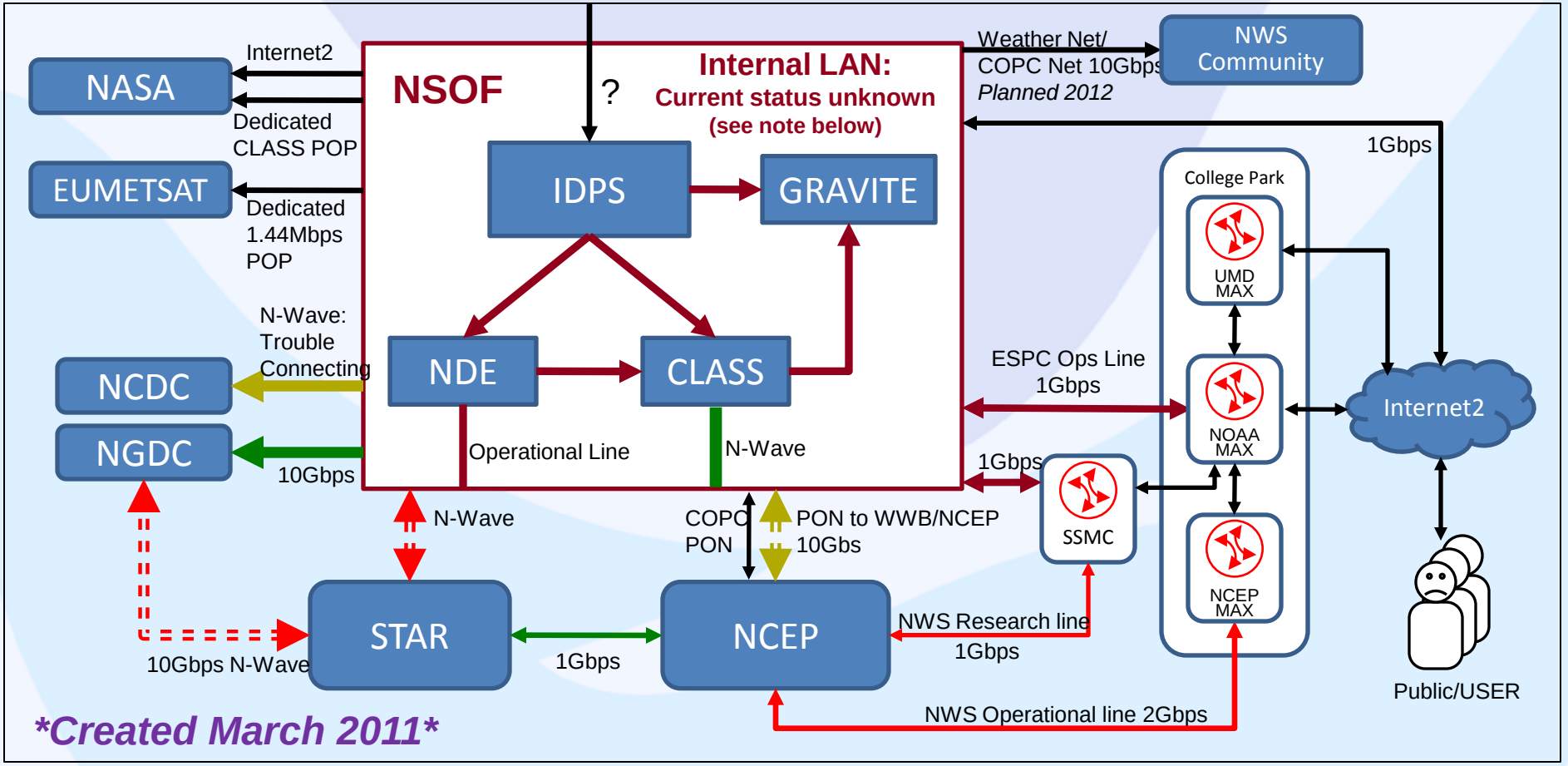
STAR Data Management and Stewardship

- All STAR data is controlled through the STAR Data Management Working Group (DMWG).
  - The group manages all commonly used incoming satellite and ancillary data and the coordinating infrastructure.
  - By extension, the DMWG steers the data management goals of STAR and shapes future data endeavors.
- The STAR data manager built a repository to store all satellite data in a common location in order to reduce redundancy of scientists storing data individually, the STAR Central Data Repository (SCDR).
  - As the data steward at STAR, the data manager created a database and software to allow access to the SCDR internally.
  - Data in the SCDR is freely available to all STAR users, a data commons.
  - Only the DMWG controls what is input and removed from the SCDR.

Challenges Getting NPP data at STAR

- NPP brings many challenges in its shear volume of data.
  - NASA estimated NPP data volumes of all xDR data to be 4 TB per day.
  - This data will need to be disseminated by NOAA at this rate to many of its operational and research users, specifically STAR
- NOAA satellite networking infrastructure at both STAR and the Environmental Satellite Processing Center (ESPC) will not be able to manage projected data volumes in addition to all other data satellite data sets as of early 2011.
  - STAR may depend on ESPC networks for dissemination of NPP data as an upstream provider so both networks would need to be upgraded.
- The NOAA research network, N-wave, is set to be completed in early 2011 and will be able to transport NPP data without issue
  - STAR could get NPP data from the NOAA data centers if connected.
  - Data latencies will be much higher from the data centers, consistent with a 6 hour moratorium period imposed by the Program.

NPP Network Readiness from STAR Perspective



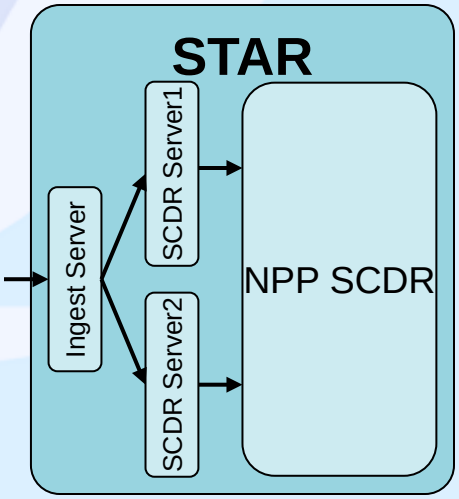
- Good, will support NPP load
- Moderate, May be able to support NPP load
- Inadequate, will not be able to support NPP load
- Highly Inadequate, will not support NPP load under any Circumstances and needs attention
- Not applicable to STAR

- Planned; Moderate Risk, will support NPP load
- Planned; High Risk, will support NPP load once in place

ESPC LAN Issues: Unknown internal load; upgrade dependent upon: funding, power, cooling, cabling, security, and weight.

Data Management in SCDR

- While attempting to finalize an NPP data source, the STAR DMWG began to build up infrastructure to meet the new data management needs.
  - Additional SCDR hardware was purchased from OSD Ground Systems funding.
  - STAR installed an additional 300 TB of NetApp storage to support NPP.
  - Servers to move the data were installed as the front end of SCDR.
  - Internal networks were bolstered to 2 GB from 1 GB to accommodate moving the data.



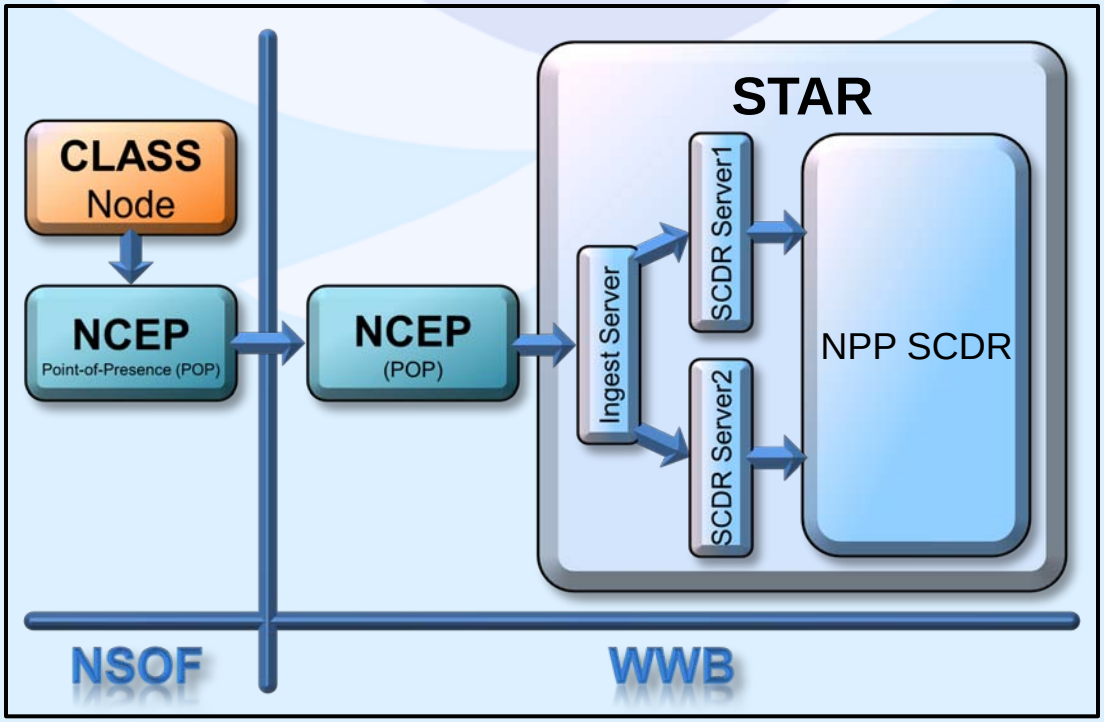
- The STAR data manager updated existing SCDR software to fully utilize all repository storage space.
- STAR set an internal goal to complete all updates by NPP launch in late October.

NPP Data Source and Networking

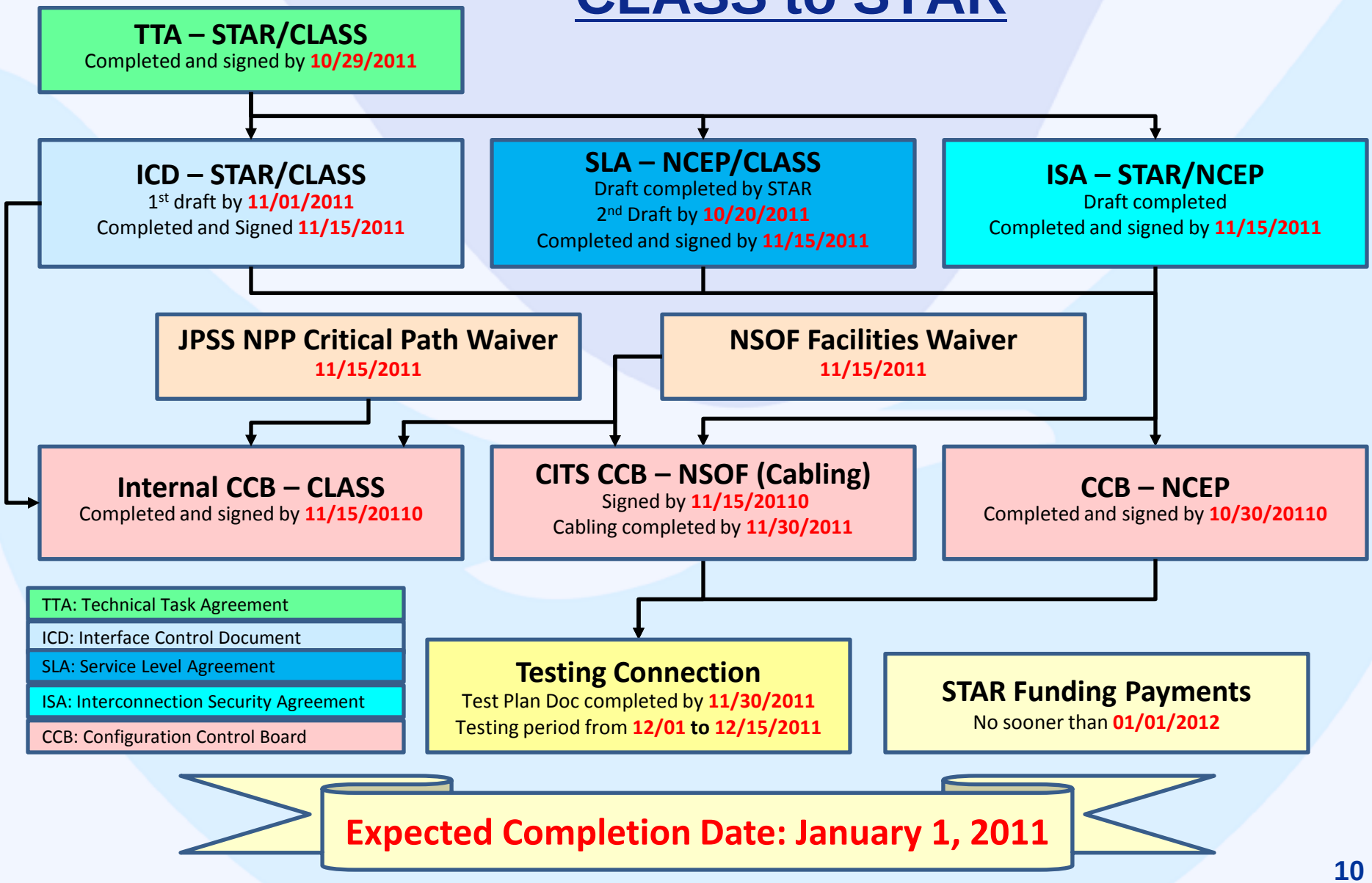
- Connecting STAR to N-wave to get NPP data from the NOAA data centers fell through.
  - HPCC funding proposal was not awarded and no internal funds could be diverted to this solution.
  - STAR would have to wait for more than a year until the move to College Park as the new building (NCWCP) would be on N-wave.
- The National Center for Environmental Prediction (NCEP) installed an operational 10 GB direct connection network to the National Satellite Operations Facility (NSOF) in August 2011 to directly obtain operational data.
  - NCEP will allow STAR to leverage this network to get NPP cal/val data on a secured segment of the line.
  - A data provider for STAR had yet to be identified, but so long as the data source were inside of NSOF, STAR could use the NCEP network.

NPP Data Provider—CLASS

- In October 2011, through negotiations, the NOAA Comprehensive Large Array-data Stewardship System (CLASS) agreed to be provide consistent data to STAR.
  - CLASS committed to a readiness date of January 1, 2012 to operationalize an NPP data flow over the NCEP network to STAR.
  - Without additional guidance, CLASS and STAR began working to put this data solution into place.



Roadmap of Tasks for Connectivity from CLASS to STAR

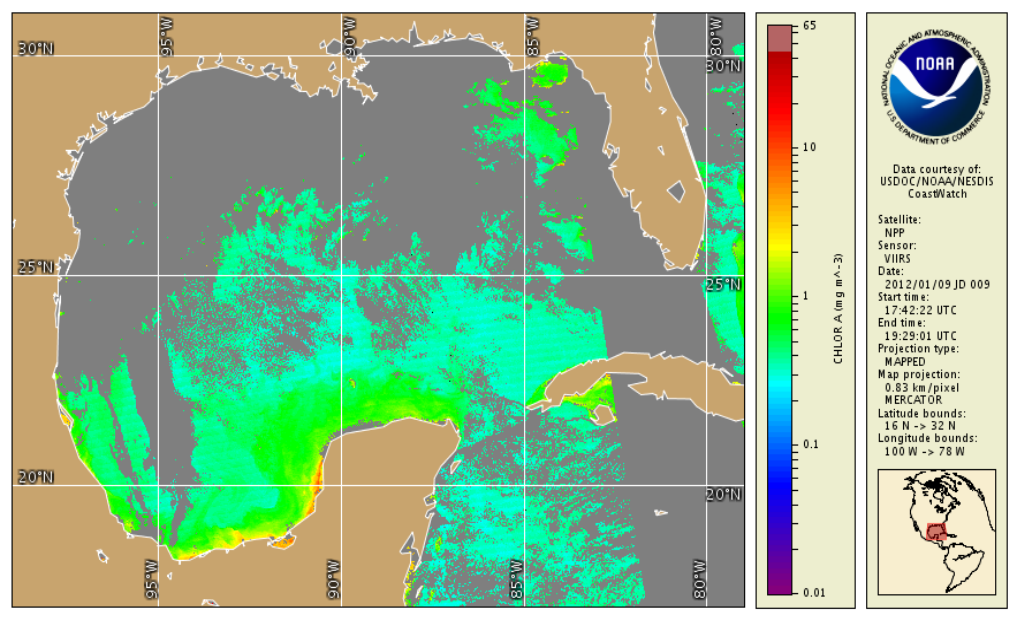


Finalizing the NPP Data Solution from CLASS

- The documentation milestones were almost all met within one week period of the proposed completion date.
  - The outstanding items for completion were the ISA documents as the scope and requirements kept changing.
- Waivers to connect the NCEP network to CLASS were a separate challenge
  - The JPSS Program Office approved CLASS system changes in one week.
  - CLASS repurposed a 10 GB server to the STAR connection, meaning no NSOF facilities waiver was required.
  - The CITS Change Control Board request to complete cabling between the NCEP and CLASS nodes took almost a month.
- The connection was fully in place and initially tested on January 17, 2012 but is not yet operational.

Continuing Challenges

- The data solution cannot be operationalized without the approval of the STAR-CLASS Interconnection Security Agreement (ISA).
  - Awaiting NOAA/NESDIS Office of the Chief Information Officer (OCIO) approval to allow the NESDIS Authorizing Officials to sign the agreement.
  - As most NPP instruments are now flowing data STAR scientists will have to continue gleaning it from ad hoc sources at CLASS and GRAVITE.
- Due to data needs, STAR is will need to operationalize SCDR with minimal testing.
- Remote access of the SCDR by STAR Cooperative Institutes and personnel will have to ride on Virtual Private Network (VPN) over the internet.



Summary of Work

- 2011 presented great challenges in disseminating NPP data to NOAA stakeholders, especially the cal/val and science monitoring teams at STAR.
- After scoping the infrastructure and data management needs, it took more than half a year to find and implement solutions to get NPP data to STAR.
- A vetted network connection and subscription was put into place in two months by diligent work at STAR and CLASS.
- The connection needs to still awaits ISA approval by NESDIS management to finalize the solution.

Through a coordinated effort of scoping data requirements and data management needs, STAR has successfully planned and implemented a solution to obtain NPP to support the Center's near and long-term goals.

Acknowledgements

Thank you to the those from STAR, CLASS, the Office of Systems Development (OSD) and the Office of Satellite and Product Operations (OSPO) who have supported my efforts therein with their time, expertise and advice.

From STAR: *Matt Jochum, Joe Brust, Laurie Rokke, Kevin Garrett, Thomas King, Walter Wolf and the entire STAR Data Management Working Group.*

From CLASS: *Kern Witcher, Dean Carter, Steve Milinovich, Al Bibbero, Robert Romero, Constantino Cremidis and Scott Koger.*

From NCEP: *Luis Cano, Cameron Shelton, Ben Kyger, Rene Rodriguez .*

From OSD: *Thomas Schott, Kevin Berberich and the NPP Data Exploitation (NDE) Team.*

From OSPO: *Russell Dyson, Will O'Dell and Andre Hammond.*

Please contact me with any inquires at

Kathryn.Shontz@noaa.gov

A copy of all documentation mentioned in this presentation can be provided upon request.