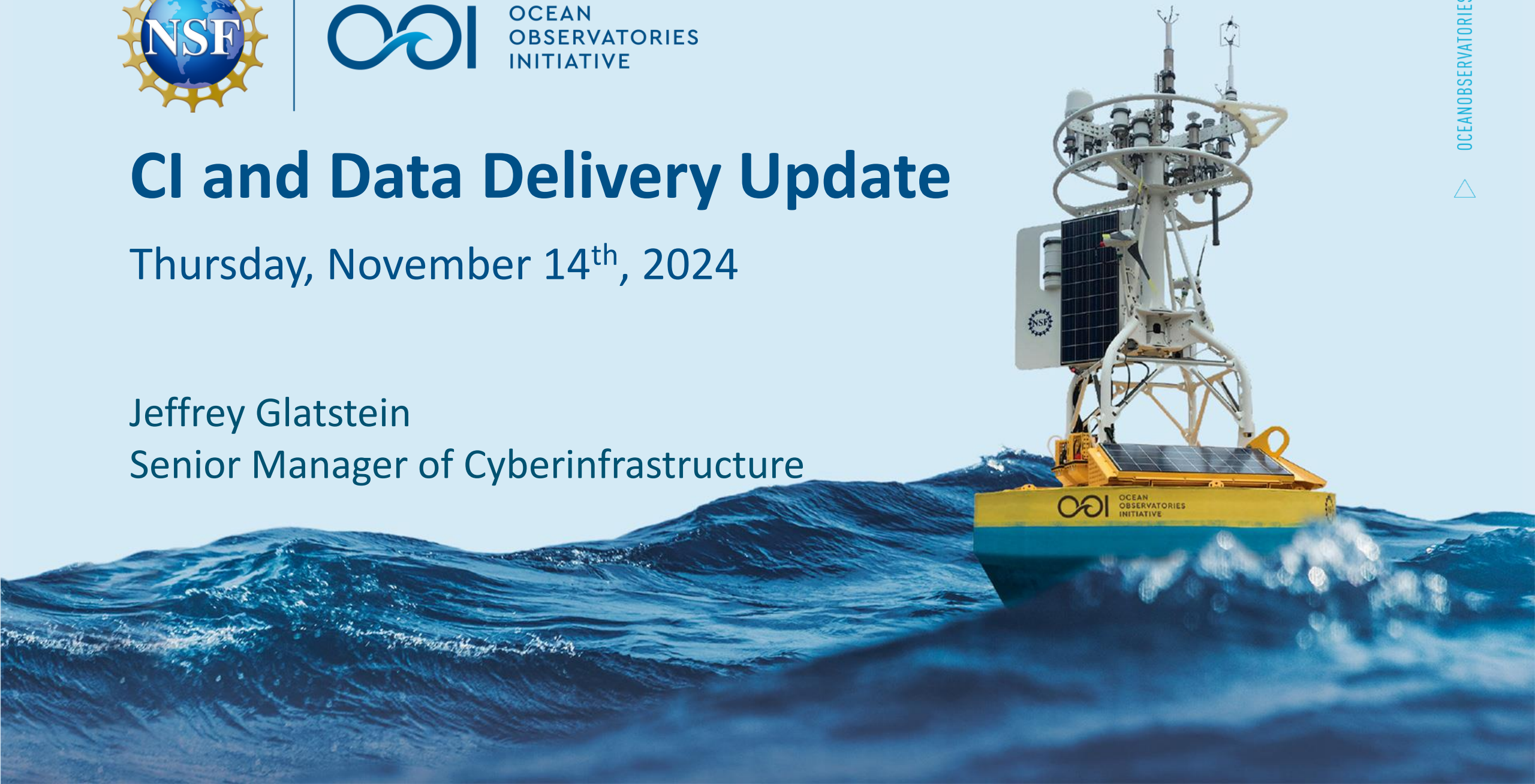
OCEAN OBSERVATORIES INITIATIVE

CI and Data Delivery Update

Thursday, November 14th, 2024

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure



OCEANOBSERVATORIES.ORG

Agenda

- PYVI Recap
- PYVII Notable Projects
- Questions

PYVI Work Highlights

- ✓ Recruit Senior Software Engineer
- ✓ Recruit Chief Security Officer – physical or virtual TBD
  - Engaged with CyberSecOp
- ○ QA/QC
  - ✓ Continue to work Data Quality and FAIR Redmine tickets
  - ✓ Build data quality report
  - ✓ Support current test logic and testing efforts
  - • Design QA/QC dashboard as one stop portal for test support
    - Parameter maintenance UI 80% completed
- ○ Continue to support pre-load changes in support of CF compliance
  - Preload changes have been migrated, but work remains for this effort
- ○ Python 2 –> Python 3 upgrade (multi year project)
  - Stream Engine and ION functions completed but not migrated.
- ○ Implement new Document Management tool

PYVI Work Highlights - continued

- ✓ Data Center Technical refresh
  - ✓ Databases
    - Upgrade of uFrame software stack was not required to complete task
- ○ System Dashboard – Phase 1
  - Architecture and product selected (Grafana). PYVII deliverable.
- ✓ Support of Pioneer move
  - ✓ New parsers
  - ✓ Update preload database
  - ✓ Migration to Data Explorer
- ○ Digital Object Identifiers
  - ✓ Sites entered and set to ‘findable’ in DataCite database
  - • Build out oceanobservatories.org support pages
  - • Implement PID logic (PYVII finish)
  - • Build look up function for DOI/PID combination
- ✓ Software administration tuning
  - Improved scripting and trained developer to perform software migrations

PYVI Work Highlights - continued

- Data Explorer
  - Display DOIs
  - Integrate HYDBB data and video into search and visualization
    - ✓Video data is completed
  - ✓Annotation service
  - ✓Operational and performance tuning
    - Significantly improved data ingestion and reload
  - Continue to implement improvements for high resolution data sets
    - Significant progress with more work to do
  - ✓QC results visualization with HITL annotations
  - Streamline system metadata management
    - Ongoing
  - ✓Integration of JupyterHub
    - Links have been added to front page
  - Research visualization of full-resolution three-dimensional data sets
  - Cross training of resources across CI
    - Progress made
  - ✓Support Data Center server virtualization

PYVI Work Highlights - continued

- Strategic
  - ✓Start research into data download redesign (replacement of ERDDAP) – multi year project
    - Progress made – rough architecture determined.
  - User interface to manage data reloads into Data Explorer
  - ✓Continue cloud research
    - Consider Jupyter HUB a private cloud.
- Impactful unplanned activities
  - ✓Adjustment of initial Data Explorer plot parameters to reduce out of range parameter impacts
    - Removed min/max envelope and climatology high/low as default plotting features. Reduced the impact to auto scaling across a large portion of the plotted parameters.
  - Refactoring of data reloads into Data Explorer to include data deletion
    - Prior to change data reloads only applied new data or updated existing data. This left data that no longer existed due to upstream corrections still displaying. QARTOD results were the test case which highlighted this.
    - Applied more automation and logging for research
    - Update is 90% completed. > 4 month impact
  - Cabled data fix for incorrect deployment number
    - Early data ingestion (pre 2018) have deployment numbers greater than 0 in the database. This was identified as the cause of data gaps.
    - Three fix cases exist – records where deployment is not zero but is the only record, records where deployment is 0 and > 0 but all other data is equal and records where deployment is 0 and > 0 but other data does not equal each other.
    - Enhanced StreamEngine to deliver raw data without applying any calibrations or logic.
    - Enhanced EDEX to delete data by deployment
    - Coding completed. Fix is starting. >5 month impact

PYVII Notable Projects

- Implement user statistics logging in Data Explorer
- Integrate visualization of IFCB dashboard into Data Explorer
- Rebuild ERDDAP back-end to improve stability and performance
- Data visualization and ingestion improvements for profilers (OPTAA)
- Python 2 -> 3 upgrade – data ingestion and MI instrument for HYDBB
- Cybersecurity implementation of CIS control sets 1 and 2
- Build out Jupyter Hub ecosystem – example sets for data connectivity both internal and external, build method of notebook discovery and look to increase Data Explorer connectivity.
- Data Quality – support MIO testing, develop Gap and Timing test logic
- Construct Engineering Data management roadmap
- Build more user-friendly raw data server search interface
- Removal of unused features and functions in OOINET
- Explore use of AI in OOI and how to prepare data for users

OCEAN OBSERVATORIES INITIATIVE

Questions?