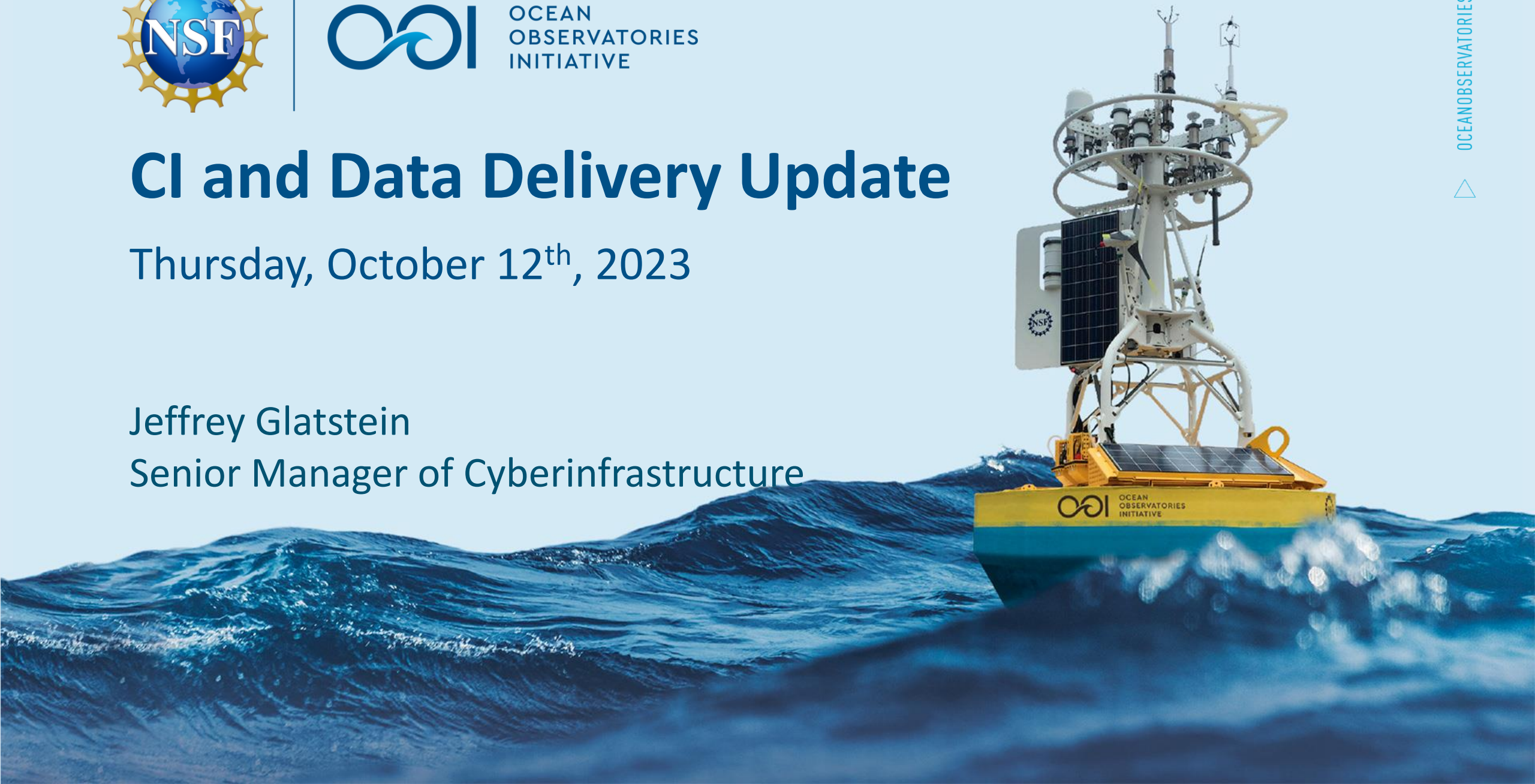
OCEAN OBSERVATORIES INITIATIVE

# CI and Data Delivery Update

Thursday, October 12th, 2023

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure




OCEANOBSERVATORIES.ORG

# Agenda

- PYV Highlights Results
- PYVI Work Plan Highlights
- Questions


OCEANOBSERVATORIES.ORG

NSF | OCEAN OBSERVATORIES INITIATIVE

# PYV Work Plan Highlights – Presented 10/2022

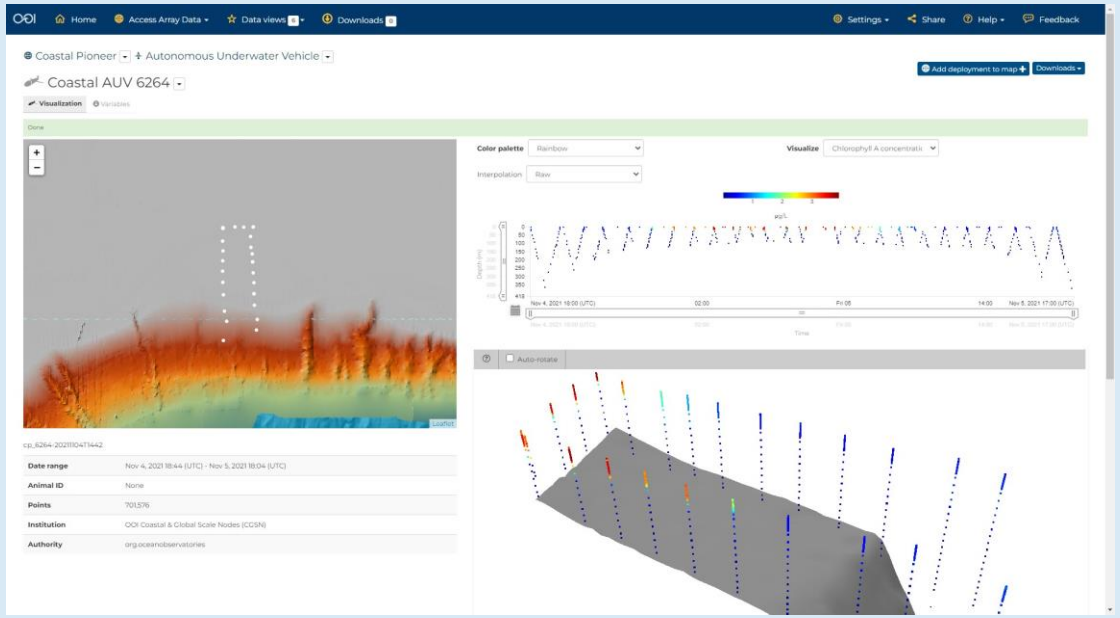
- Stream Engine re-architecture
  - Upgrade to Python 3 (SE Code and all ION functions)
  - 30+ requirements – Reporting across reference designators, .zarr file support, multi-level co-located instrument data
  - Data request management – load balancing, request management routes to cancel requests
  - DOI/PID implementation

❖Requirements completed and reviewed
❖Coding to start in early PYVI

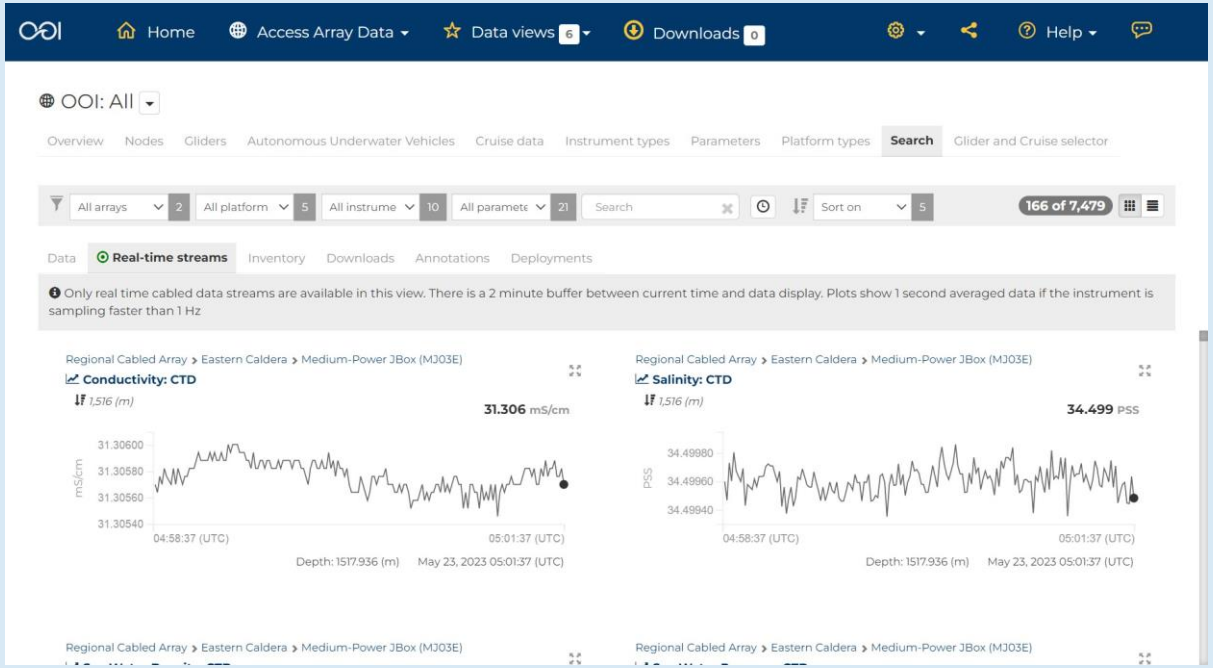
# PYV Work Plan Highlights – cont.

- Data Explorer
  - ➢Completion of full resolution data visualization
    - ➢More data has been loaded which requires tuning at each load. Instrument by instrument approach.
  - ➢Expansion of media server to include HD video, Hydrophone and streaming data
    - ✓Streaming data added
    - ➢Video compression and server put in place
    - ➢Hydrophone had many processing updates but were not yet added to Data Explorer
  - Data Explorer operational training to OOI development and operational resources
    - ✓Started – multi year effort
  - Further reingestion automation and reporting
    - ✓Further error detection added
    - ✓Dashboard for internal use available
    - ➢V&V (Verify and Validate) script in process – Verification in testing
  - ✓ZPLS and AUV data availability
  - ➢Addition of remaining scientific data

# PYV Work Plan Highlights





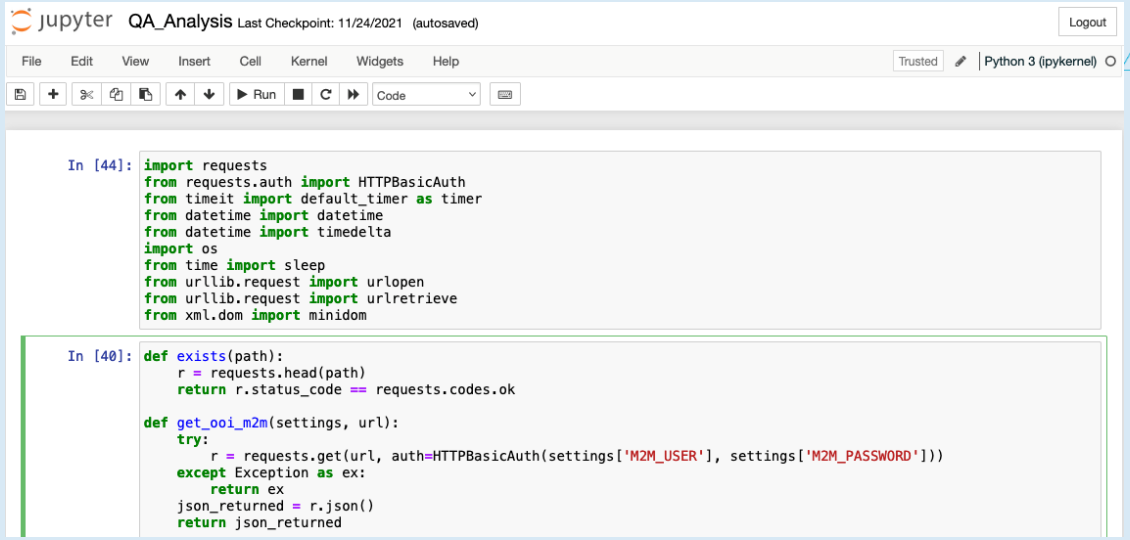




OCEANOBSERVATORIES.ORG

NSF OCEAN OBSERVATORIES INITIATIVE

# PYV Work Plan Highlights – cont.

- Compute in place - JupyterHub beta release
  - ✓Used for 2023 Summer School
  - ✓In production with sign-up required
  - ➢Link to Data Explorer




➢ Asset management – Roundabout development

# PYV Work Plan Highlights – cont.

- Data Accuracy and FAIR
  - ✓Continue to target data quality tickets
  - ✓Continue QARTOD support and development of tests and tools
  - ✓Continue to support preload database analysis and adjustments
    - ✓Mostly targeted changes
  - ✓Continue FAIR data standards tuning (JupyterHUB, Preload database work)
- Performance
  - ➢Query performance analysis
    - ➢Initial analysis done on recurring queries that may not need to execute as often
  - ✓Integration of new processing and storage resources
- Operational
  - ➢Cloud storage transfer to TACC
  - ✓NCEI data archival
  - ➢Dev-ops, monitoring and improved efficiency of releases
  - ❖Database replication
  - ❖Disaster recovery scenario exercises

# PYV Work Plan Highlights – cont.

- Strategic
  - ERDDAP tuning and replacement evaluation
    - ➢Tuning has continued
    - ❖Replacement evaluation moved to 2.5
  - ✓Deliver Digital Object Identifiers (DOI) recommendations for policy and approach
    - ✓First DOI issued
    - ✓Approach and architecture designed and presented
    - ✓DataCite configuration completed
    - ✓Data mock-ups completed
  - ✓Analysis of Alternatives for Alfresco, Confluence and Jira
  - ❖Evaluate options to reduce the Cassandra/PostgreSQL database footprint
  - ✓Continued cloud analysis
    - ✓3rd Party analysis with cost estimates and required efforts completed
    - ✓Outlines cloud is still currently more costly than on-prem, but the gap is closing

# PYVI Work Highlights

- Recruit Senior Software Engineer
- Recruit Chief Security Officer – physical or virtual TBD
- QA/QC
  - Continue to work Data Quality and FAIR Redmine tickets
  - Develop data quality tests based on MIO requirements (e.g. gap and timing)
  - Build data quality report
  - Design QA/QC dashboard as one stop portal for test support
- Continue to support pre-load changes in support of CF compliance
- Python 2 – Python 3 upgrade (multi year project)
  - Stream Engine and ION functions
  - Ingest engine and Parsers
  - Port Agents (MI Instrument)
- Implement new Document Management tool
- Data Center Technical refresh
  - Databases
  - Upgrade of uFrame software stack TBD
- System Dashboard – Beta
  - Data Explorer content with last loaded information
  - Data ingestion status
  - Report of data location
  - Data retrieval queues
  - System messages

# PYVI Work Highlights

- Support of Pioneer move
  - New parsers
  - Update preload database
  - Migration to Data Explorer
- Digital Object Identifiers
  - Build out oceanobservatories.org support pages
  - Build look up function for DOI/PID combination
  - Implement PID logic (PYVII finish)
- Software administration tuning
- Data Explorer
  - Display DOIs
  - Integrate HYDBB data and video into search and visualization
  - Annotation service
  - Operational and performance tuning
  - Continue to implement improvements for high resolution data sets
  - QC results visualization with HITL annotations
  - Streamline system metadata management
  - Integration of JupyterHub
  - Research visualization of full-resolution three-dimensional data sets
  - Cross training of resources across CI
  - Support Data Center server virtualization

# PYVI Work Highlights

- Strategic
  - Start research into data download redesign (replacement of ERDDAP) – multi year project
  - User interface to manage data reloads into Data Explorer
  - Continue cloud research

OCEAN OBSERVATORIES INITIATIVE

# Questions?