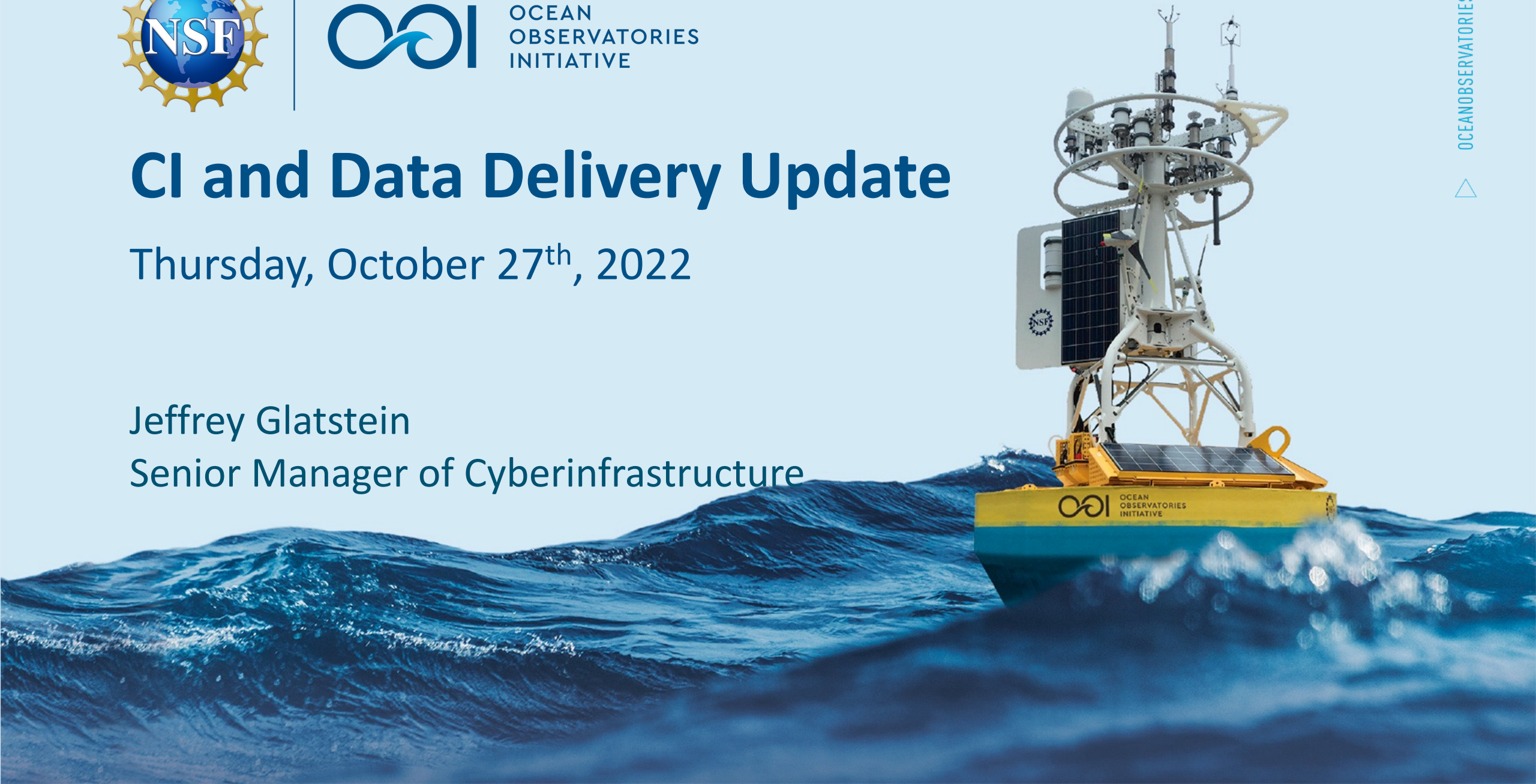
OCEAN OBSERVATORIES INITIATIVE

CI and Data Delivery Update

Thursday, October 27th, 2022

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure



OCEANOBSERVATORIES.ORG

Agenda

- PYIV Work Plan
- PYIV Highlights
- PYV Work Plan Highlights
- Questions

OCEANOBSERVATORIES.ORG

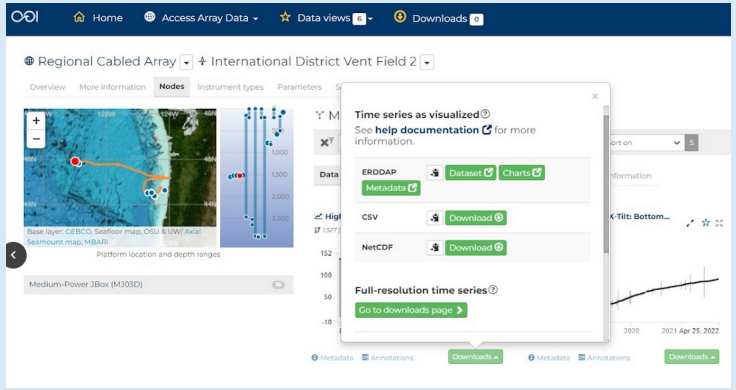
OCEAN OBSERVATORIES INITIATIVE

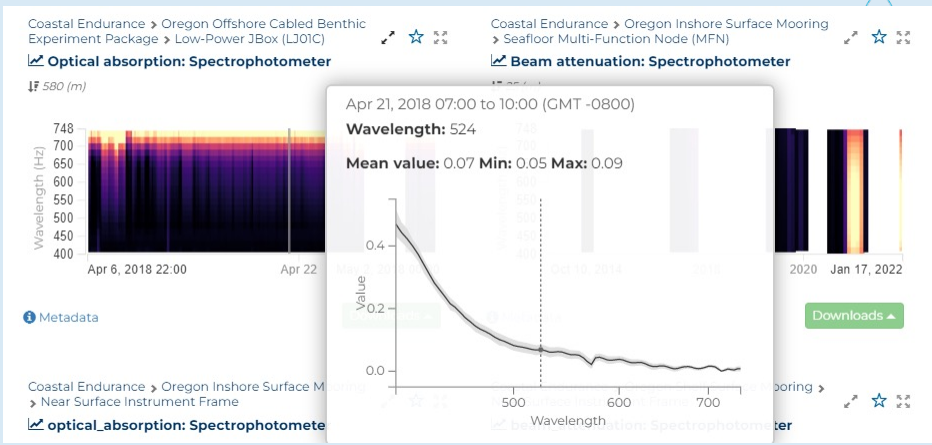
PYIV Work Plan – presented 12/07/2021

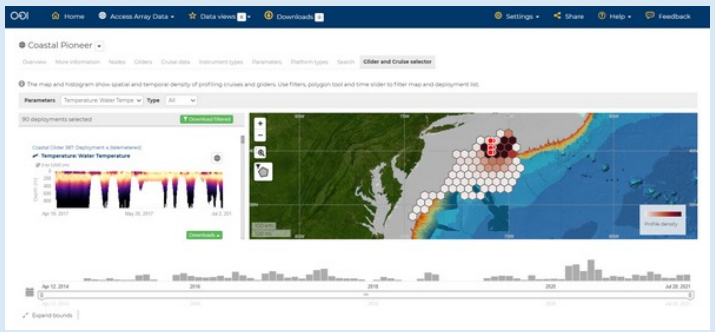
- Data Explorer v1.3 and v1.4
  - Visualization of multi-dimensional plot data (e.g. spectral)
  - Tuning of data ingest re-load
  - Ingest ZPLS* echopype generated data streams and produce echograms
  - Media server assets – video and photos
- Expansion of Jupyter notebook access
- Stream Engine modernization
- Instrument centric approach on ticket resolution
  - Pressure/Depth
  - PHSEN
  - Dissolved Oxygen
- QARTOD support
  - Climatology and gross range implementation
  - Gap and timing requirements gathering
- Complete the Roundabout connectivity to uFrame
- Pre-load and NetCDF compliance improvements continue

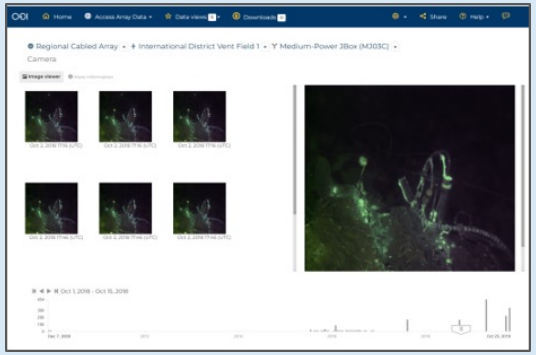
PYIV Work Plan Highlights

- Data Explorer v1.3 and v1.4
  - ✓Visualization of multi-dimensional plot data (e.g. spectral)
  - ✓Tuning of data ingest re-load
  - ✓Ingest ZPLS* echopype generated data streams and produce echograms
  - ✓Ingest AUV data
  - ➢Media server assets – video and photos
  - ✓PI data sites findable on landing page
  - ✓Discrete data searchable and available
  - ✓Improved download interface language







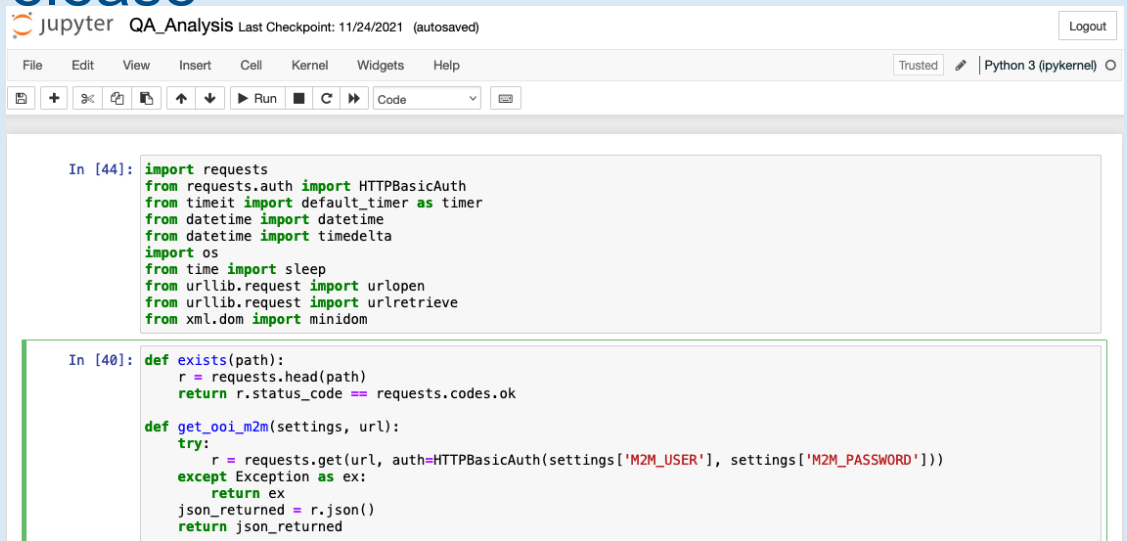


OCEANOBSERVATORIES.ORG

NSF | OCEAN OBSERVATORIES INITIATIVE

PYIV Work Plan Highlights

- ✓ Expansion of Jupyter notebook access
  - Completed POC for single notebook user on CI server
  - Completed POC for multi-user tiny Hub set-up
  - Applied STAC logic for organization
  - Prepared and presented example notebooks
  - Set environment for next step of Beta release
  - More secure environment



OCEANOBSERVATORIES.ORG

NSF | OCEAN OBSERVATORIES INITIATIVE

PYIV Work Plan Highlight

- Stream Engine modernization
  - Not started in favor of continuing focus on data quality efforts
- ✓ Instrument centric approach on ticket resolution
  - ✓Pressure/Depth
  - ✓PHSEN
  - ✓Dissolved Oxygen
- ➢ QARTOD support
  - ✓Climatology and gross range implementation
  - ➢Gap and timing requirements gathering.
    - ➢Far down the path but more discussion required
- ➢ Complete the Roundabout connectivity to uFrame
  - ➢Connectivity testing started but not completed
- ✓ Pre-load and NetCDF compliance improvements continue

PYV Work Plan Highlights

- 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
  - Data request management
- Data Explorer
  - Completion of full resolution data visualization
  - Expansion of media server to include HD video, Hydrophone and streaming data
  - Data Explorer operational training to OOI development and operational resources
  - Further reingestion automation and reporting
  - ZPLS and AUV data availability
  - Addition of remaining scientific data
- Compute in place - Jupyter Hub beta release
- Asset management – Roundabout development

PYV Work Plan Highlights

- Data Accuracy and FAIR
  - Continue to target data quality tickets
  - Continue QARTOD support and development of test and tools
  - Continue to support preload database analysis and adjustments
  - Continue FAIR data standards tuning (Jupyter HUB, Preload database work)
- Performance
  - Query performance analysis
  - 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

- Strategic
  - ERDDAP tuning and replacement evaluation
  - Deliver Digital Object Identifiers (DOI) recommendations for policy and approach
  - Analysis of Alternatives for Alfresco, Confluence and Jira
  - Evaluate options to reduce the Cassandra/PostgreSQL database footprint
  - Continued cloud analysis

OCEAN
OBSERVATORIES
INITIATIVE

Questions?