OCEAN OBSERVATORIES INITIATIVE

CI Data Delivery Update

Tuesday, May 6th, 2025

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure

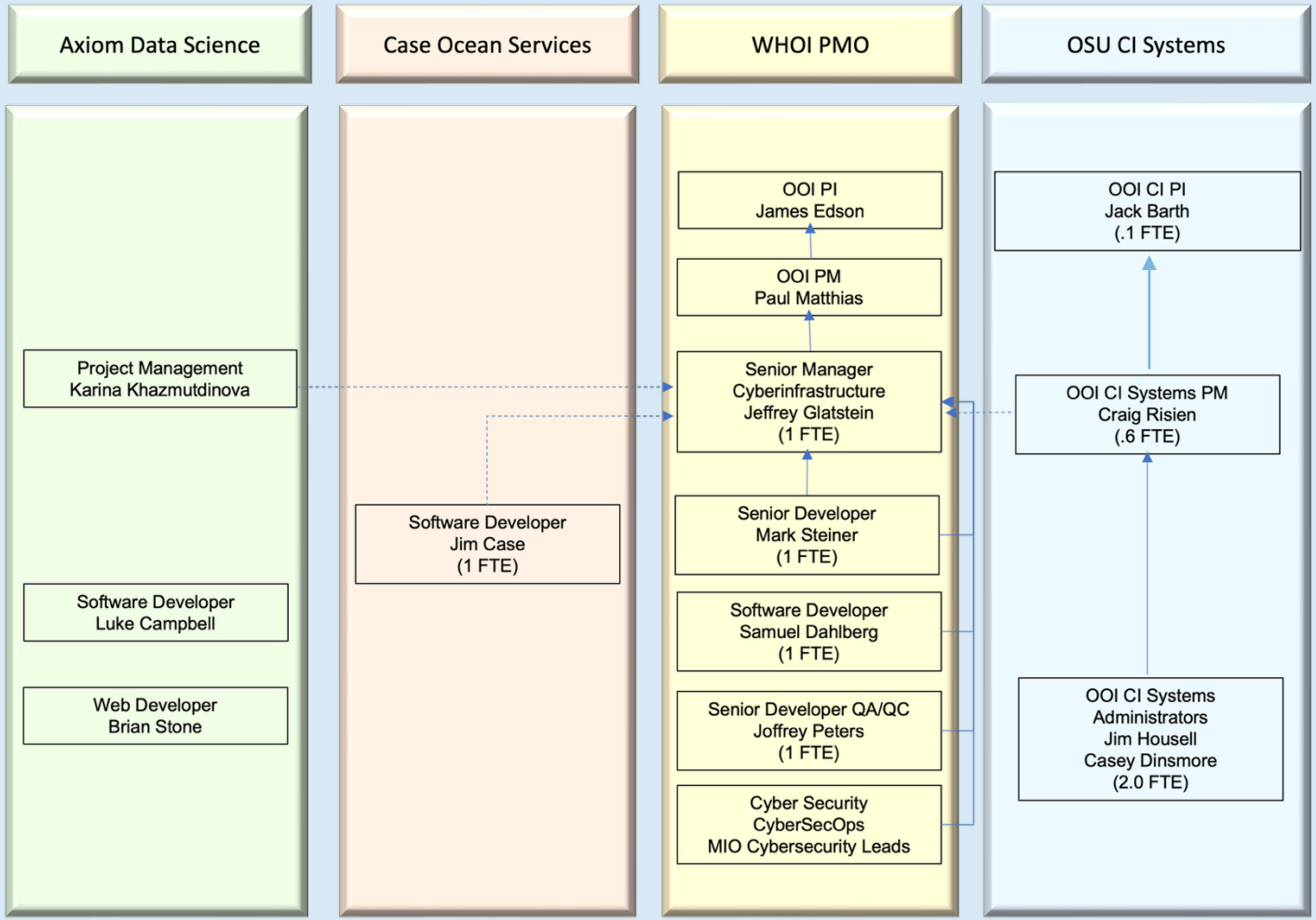
OCEANOBSERVATORIES.ORG

Agenda

- CI Team
- PYVII Project Status
- Questions

CI Team

- Senior Manager of Cyberinfrastructure and Data Delivery Manager (PMO) - responsible for all aspects of the OOI Cyber Infrastructure (strategy, budget, and execution), data delivery (including UX), and execution of a QA/QC program.
- CI Systems Project Manager (OSU) – responsible for day-to-day operations, including prioritization of operational tasks, management of Systems Administrators, budgetary execution for purchases and renewals, executing on strategic priorities, and development and submission of required reports.
- Systems Administrators (OSU) – responsible for the monitoring and maintenance of the OOI CI hardware and network infrastructure.
- Lead Software Engineer (PMO) – responsible for uFrame and data ingestion components and tasked with reviewing other developer's designs and code.
- Software Developer(s) PMO – Concentrate on data quality and DevOps tasks.
- Software Developer (Case Ocean Services) – responsible for maintaining and retiring the legacy Data Portal, web services supporting Data Explorer, multi-media processing and asset metadata delivery.
- Project Manager (Axiom Data Sciences) – responsible for coordination and management of Axiom resources developing the Data Explorer tool.
- Software Developer (Axiom Data Sciences) – responsible for data ingestion and interface processes into the Data Explorer tool.
- Web Developer (Axiom Data Sciences) – responsible for the UI for the Data Explorer tool.
- Cyber Security Leads – direct the Cyber security efforts across the OOI program in conjunction with PMO Developers, OSU Systems Admins and MIO Security Leads. CyberSecOps will act as a vCISO.



PYVII Work Highlights

- Data Explorer
  - Implement usage statistics in support of reporting on user impact
    - Not yet started and possibly won't this project year. Usage impact is currently analyzed using log data.
  - Integrate visualization of Image Flow Cytobot (IFCB) data
    - 60% complete. This will be a two-pronged approach with links to the IFCB dashboard being built into the instrument search and the header data being integrated into search later.
  - Rebuild ERDDAP back-end to improve stability and performance
    - This is a multi-year effort with a small time dedicated to starting this program year. It has not started yet.
    - Work has been done to simplify how data is stored and ingested. This will ultimately help the ERDDAP project.
  - Gather requirements to build roadmap for implementation of next-gen portal architecture
    - This is an internal project to Axiom. Good progress has been made.
  - OPTAA data visualization improvement
    - A working group was formed last project year. Work was put on hold due to unplanned activities impacting the required development resources.
    - Plan is to reconvene the working group this Spring.

PYVII Work Highlights

- Data Explorer Continued
  - Improve data processing and visualization for profiler data, wave data or curtain plot resolution
    - Impacted by unplanned activities. Many of the work products from those activities are applicable to this effort. In particular, the data processing of difficult data sets like profilers.
- Data Explorer Unplanned Activities
  - Refactor of data reloads from OOI databases into Data Explorer databases
    - Prior to this effort, there was the possibility of old data remaining after a reload of data in Data Explorer.
    - This project has made the data, in particular automated QC test results, more accurate.
  - Refactor of ingestion for profiler data
    - This project allows data from deep profilers to be visualized and made the ingestion system more stable.
  - Refactor of ingestion for large data sets (ADCP and VADCP)
    - Parts of this effort will be used in future projects: ERDDAP rewrite, visualization for OPTAA, system stability
    - This project is 99% completed. Once finished, all ADCPs and VACPs will be visualized in Data Explorer
    - Switched the underlying file structure from NetCDF to .ZARR.

PYVII Work Highlights

- Operational Improvement
  - Python upgrade of ingestion, parser code and MI instrument (HYDBB) to version 3.x
    - This phase of the Python upgrade has not started. Stream Engine and ION code has been upgraded but not migrated to production. This was impacted by unplanned activities.
  - Design and implement user interface for MIO scheduling of data reloads from OOI databases
    - Impacted by unplanned activities, particularly data reloads. This effort may not be rescheduled.
  - Separation of monitoring metrics and query analysis from production database instances
    - Project has been cancelled as other projects and unplanned activities have a higher priority.
  - Upgrade to Cassandra version 5 including Java and Spring drivers
    - Research for this is complete. This will be pushed to next project year as an upgrade of EDEX is a prerequisite.
  - Gain operational efficiencies by adding further automation to Software and Asset Management scripts and processes
    - Completed. This will be an ongoing iterative approach. The biggest change has been completed. The next iteration will be moving software packaging to Git.

PYVII Work Highlights

- Community Experience
  - Example notebooks for processing video, hydrophone, ZPLS and general OOI data search
    - Much of this work has been completed. The method of sharing needs to be worked.
  - Set-up STAC (Spatio Temporal Asset Catalog) in Jupyter notebook
    - This effort has not been started.
  - Build data connectors in Jupyter Hub to other data sources like NEON and ONC
    - This effort will start in June with the new CI Compass Fellows.
  - Machine learning QA/QC code for hydrophone, video and digital stills
    - This effort will continue in June with the new CI Compass Fellows.
  - QA/QC support
    - This is an ongoing effort with the implementation of new test parameters and code development.
  - Construct Engineering data management roadmap
    - In progress. Usage statistics have been gathered and analyzed.
  - QA/QC status information available in Data explorer
    - This will be a link to the OOI QA/QC status page in oceanobservatories.org

PYVII Work Highlights

- Strategic
  - Develop method to search raw data by common tags
    - Not started. May not be prioritized this project year.
  - Build API to GraphQL interface to utilize data already produced by Stream Engine in real-time
    - Cancelled. Objectives were met through database connectivity.
  - Removal of unused features in OOINET
    - Not started.
  - Re-architect M2M to include intelligent data retrieval and versioning of data sets.
    - Not started. May not be prioritized this project year.
  - System status dashboard
    - Tool selection is complete, and a proof of concept (POC) has been assembled.
  - Explore AI and how the OOI can better prepare data for future use through the implementation of a proof of concept (POC) project.
    - Three questions have been proposed and review with external AI expertise is being sought.

PYVII Work Highlights

- Prior AWP and unplanned efforts
  - BOTPT tide table implementation
    - Prior AWP - completed. This effort introduces the concept of virtual streams being stored in Cassandra.
  - Co-located sensor data missing data issue
    - Unplanned effort - completed. This fix will align the co-located data retrieved with the stream method of the original data (e.g. Telemetered vs Recovered).
  - VADCP-A existing code replaced with VADCP-B new code
    - Unplanned effort – completed. This enhancement allows the new model of VADCP to deliver data.
  - Prawler (MAB) parser development
    - Prior AWP item – completed. This effort allows the recovered data to be processed into the databases.
  - SeaPHOx parser
    - Unplanned effort – completed. This enhancement allows the new instrument to deliver data.
  - Cabled Data deployment number fix
    - Unplanned effort - in progress. Data ingestion early in the program's history has the wrong deployment number. This is a significant effort and will remove data with the wrong deployment data and correct those with the wrong deployment.

OCEAN OBSERVATORIES INITIATIVE

Questions?