



OCEAN
OBSERVATORIES
INITIATIVE

OOI Cyberinfrastructure AWP Review

Tuesday, May 7th, 2024

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure



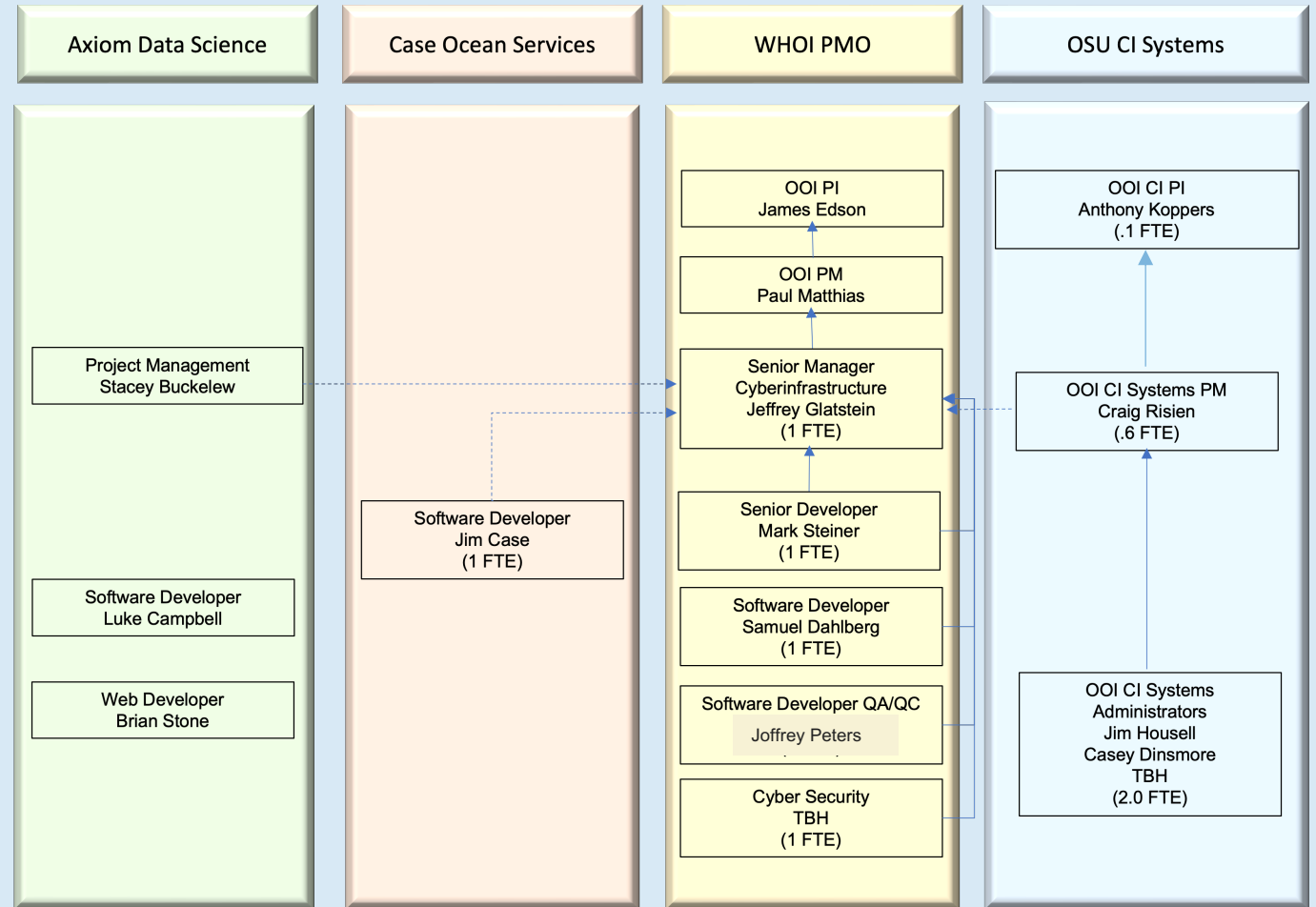
Agenda

- ❖ Review PYVII Objectives
- ❖ Topics of Interest to CI
- ❖ Questions



CI Resources

- 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 Lead (CISO) – direct the Cyber security efforts across the OOI program in conjunction with PMO Developers, OSU Systems Admins and MIO Security Leads.



PYVII Tasks

Task	Impact	Resource	Planned Delivery
Implement usage statistics logging in Data Explorer to support reporting on user impact.	This will help with user experience as use patterns will emerge, as well as reporting out overall impact to the user community.	Axiom	September 2025
Integrate visualization of IFCB data into Data Explorer	This will improve the FAIR position for this data and improve user experience by eliminating or streamlining access to this data currently viewable outside of Data Explorer. This addresses, in part, the DSC long-term recommendation to improve how users find data.	Axiom	September 2025
Rebuild ERDDAP back-end to improve stability and performance.	This will improve the FAIR position for large datasets and improve user experience by enabling ERDDAP to serve data sets that it cannot currently do because of size. These data sets are available to download from THREDDS. This addresses DSC long-term recommendation for replacing ERDDAP with scalable solutions, while also addressing user concerns of changes to the ERDDAP interface.	Axiom	September 2025
Gather requirements to build roadmap for implementation of next generation Data Explorer portal architecture.	This will improve user experience and implement efficiencies for better system maintenance. Research of new data collection/dissemination processes will be part of this effort. This addresses, in part, the DSC long-term recommendation to keep current on data science advance, methods and tools.	Axiom	September 2025



PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
OPTAA data visualization improvement	Improved data consumption. This effort will concentrate on profiler data. User input will be sought for this effort.	Axiom	April 2025
Improve data processing and visualization – for example profiler data, wave data or curtain plot resolution	Improved data consumption and better visuals to help users understand the data. User input will be sought for this effort. This addresses, in part, the DSC long-term recommendation to keep current on new methods of data science.	Axiom	April 2025
Planned and ad-hoc system maintenance and emergent issues	This will improve system stability and efficiency, plus provide a better user experience by resolving reported issues.	Axiom	September 2025
Python upgrade of ingestion and parser codes from version 2 to version 3.x	This will add to system stability as well as open up a new class of libraries and cybersecurity functions for future use. This is considered required maintenance.	PMO CI Developers	March 2025



PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
Python upgrade of MI instrument code for hydrophones from version 2 to version 3.x	This will add to system stability as well as open up a new class of libraries and cybersecurity functions for future use. This is considered required maintenance.	PMO CI Developers	June 2025
Design and implement a user interface for Marine Operators to schedule data reloads into Data Explorer and schedule new ingestions into the OOI databases.	This will add efficiency to the data movement processes within the OOI systems, improving timeliness of data migration and eliminating a step in the process.	PMO CI Developers	April 2025
Six software releases, one every other month	Make fixes and enhancements available to the user community	PMO CI Developers	TBA with notice
Addressing ad-hoc and continual software maintenance, emergent issues and feedback.	This will improve system stability and efficiency, plus provide a better user experience by resolving reported issues.	PMO CI Developers	September 2025



PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
Four system maintenance windows, one per quarter	System stability and maintenance	CI Systems	TBA with notice
Yearly quiet database backup	Backing up the OOI databases together once a year provides a restorable point in time which is critical to resilience from a disaster or cybersecurity event. OOI data would only need be re-ingested from that point forward saving months in recovery time.	CI Systems	October 2024
Security updates as needed with notice	These updates will be driven by external events and vendor updates to known threats. This is required to keep the OOI system safe and in best practice compliance. Notice will be provided as best as is feasible.	CI Systems	TBA with Notice
Separation of monitoring metrics and query analysis from production database instances to read only copies.	This will improve system stability and reporting of issues with minimal impact to the instances.	CI Systems	March 2025
Research and implement a patch management system	This will improve system maintenance efficiency and stability.	CI Systems	September 2025



PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
Upgrade to Cassandra version 5 including Java and Spring drivers	This will keep OOI on the latest version of database software and support, allowing OOI to use any new features and performance enhancements.	CI Systems, PMO CI Developers	September 2025
Gain operational efficiency by adding further automation to Software and Asset Management scripts and processes	This will improve system maintenance efficiency and stability.	CI Systems	September 2025
Planned and ad-hoc system maintenance and emergent issues	This will improve system stability and efficiency, plus provide a better user experience by resolving reported issues.	CI Systems	September 2025
Execute on security architecture plan developed by the collaboration of OmniSOC and OOI in PYVI, inclusive of implementation of CIS controls.	This will build upon and harden current security protocols and implement a more rigid reporting and resource management procedures. Cybersecurity is a necessary component of a healthy cyberinfrastructure. A separate timeline for this activity will be published in the yearly cybersecurity plan.	PMO CI Security	September 2025





PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
Support security architecture (e.g. scripting in support of CIS controls like Inventory)	Inventories and monitoring of them is key to a secure cyberinfrastructure. This work in conjunction with building out the security architecture is required to implement the cybersecurity roadmap.	CI Systems	September 2025
Example notebooks for processing video, hydrophone, ZPLS and search OOI data	Introduces the Jupyter Hub environment and reduces the barrier to use by providing useable methods to search and use data without first downloading. This will have a direct impact on OOI FAIR posture as it improves access, discoverability and processing. This addresses, in part, the DSC long-term recommendations to find data with cloud-based technologies and use on-prem compute offerings.	COS	May 2025
Set-up STAC (SpatioTemporal asset Catalog) in a Jupyter notebook	This will improve user experience and data discovery by providing an alternate method of searching OOI data. This addresses, in part, the DSC long-term recommendations to utilize new methods for data science and improve a user's ability to find data and explore NSF OOI datacenter on-premises compute offerings.	COS	May 2025
Build data connectors in Jupyter Hub that will help translate OOI data structures with other complimentary data sources such as ONC or NEON.	This will improve the OOI FAIR posture by improving interoperability. It will also help researchers expand data use across multiple sources. This addresses, in part, the DSC long-term recommendations to utilize new methods for data science and improve a user's ability to find data.	COS	September 2025



PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
Machine Learning QA/QC code for hydrophone, Video and Digital Stills	This will provide users of large file raw data a means to identify those files of little use and improve data quality. It will also allow for a quicker identification of issue prior to being noticed in an HITL process. This addresses, in part, the DSC long-term recommendations to utilize new methods for data science.	PMO CI Developers	July 2025
Move Flatline legacy QC code into a QARTOD style of parameter and code management.	This will improve data quality as well as maintainability of both the QC parameters and its code.	PMO CI Developers	March 2025
Develop Gap QC test.	This will improve data quality.	PMO CI Developers	August 2025
Develop Timing QC test.	This will improve data quality.	PMO CI Developers	September 2025
Work with MIOs to construct an Engineering data management roadmap inclusive of discovery and delivery requirements.	This project will set the roadmap for PYVIII and where engineering data will live. The overall effort will improve user experience and data discovery.	Axiom/CO S/PMO CI Developers	September 2025



PYVII Tasks - continued

Task	Impact	Resource	Planned Delivery
Develop method to search raw data by common tags such as instrument name.	Significantly improves user access to raw data. This addresses, in part, the DSC long-term recommendations to continue improving the ability of users to find data.	PMO CI Developers, CI Systems	July 2025
Build an API to GraphQL interface to utilize data already produced by Stream Engine in real-time.	This will provide an alternative method to retrieve current data, improving user access to data and taking BOT query loads off of StreamEngine.	COS	March 2025
Removal of unused features and functions of OOINET and redesign of Redis cache	Improves system maintenance and user experience by removing distractions and unnecessary code.	COS	September 2025
Rearchitect M2M to include intelligent data retrieval and versioning of data sets inclusive of DOI/PID across all access interfaces	Data retrieval will become more efficient by utilizing already created data sets. It will also improve OOI's FAIR position by enhancing users' ability to recreate research results through access to historical data sets.	COS, Axiom, PMO Dev	September 2025





Topics of Interest to CI

- Use and/or preparation for AI or ML with "real-life" examples
 - Curated/Analysis ready data sets – what is of interest to the community and how might they be presented
 - Training materials – what is of interest to the community and how might they be presented
 - Methods and programs to increase user engagement
 - Data “connectors” to other repositories that will yield value to the user community
 - User experience feedback and improvements
-
- Topics that help future proof data delivery... within reason





OCEAN
OBSERVATORIES
INITIATIVE

Questions?

