OCEAN OBSERVATORIES INITIATIVE

OOI Cyberinfrastructure Strategic Initiatives

Wednesday, November 13th, 2024

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure

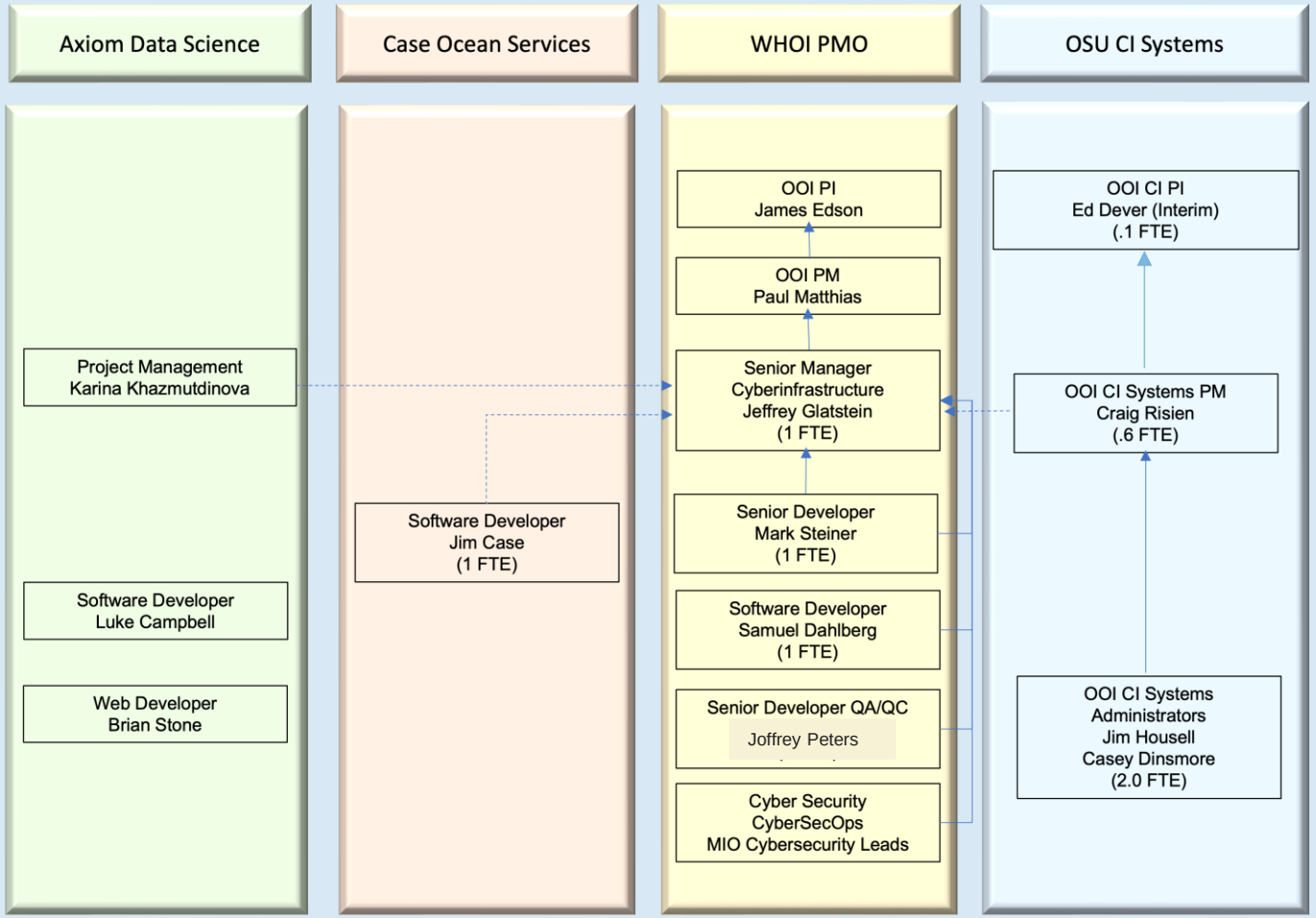
OCEANOBSERVATORIES.ORG

Agenda

- CI Team
- Creation of Filetype .zarr
- Jupyter Hub
- AI
- DOIs
- 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.



OCEANOBSERVATORIES.ORG

NSF | OOI OCEAN OBSERVATORIES INITIATIVE

Creation of Filetype - .zarr

- OOI now has the ability to create .zarr files post Stream Engine (SE) processing. SE will be enhanced to create .zarr natively.
- Need user input and testing to help tune data distribution within the file.
- How does OOI engage with the user community on this topic?
  - Email request for user input?
  - Townhall discussion?
  - Release files with ”best guess” and wait for feedback?
  - Any DSC or OOIFB members interested in providing feedback?
- Are .zarr files still of interest to the user community?

Jupyter Hub

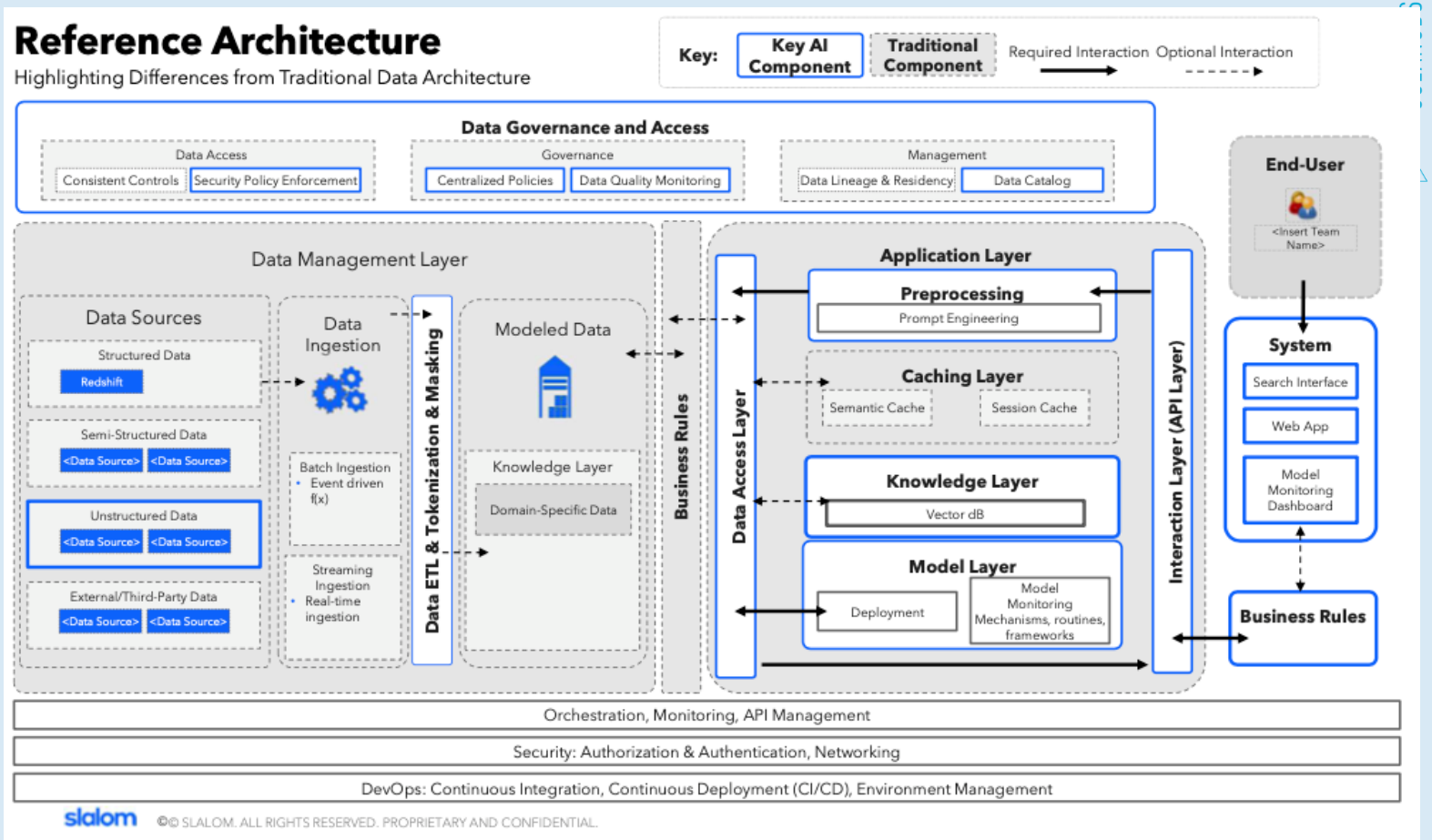
- What is the best way to store notebooks? GitHub?
- How does OOI make notebooks discoverable and distributable?
  - What would you, as users, look for to discover these notebooks?
  - How do you want to consume them?
- Should the program allow sharing of external notebooks?
  - If yes, how? Are all users eligible or should there be a vetting method? Keep in mind that all OOI Jupyter Hub users are known as access needs to be granted.
- Need to build code natively to utilize GPU. What would be a good example notebook to demonstrate this?

AI – How Can OOI Leverage the Current State

- What does OOI and the Oceanographic Community want from AI?
- Need to better understand AI and its application to science problems
  - Generative vs discriminative
  - Does one data model server all/many applications?
  - Given the speed of change in this area, how does OOI 'future' proof its efforts? Is AI mature enough?
- Continue research into AI and how OOI can prepare for the future.
  - MBARI – FathomNet - https://fathomnet.org/fathomnet/#/
  - Applied Ocean Research – AI for ocean monitoring and modelling: https://www.sciencedirect.com/special-issue/10FCZL672Q7
  - Schema.org
- Engaged with US NSF Compass CI to discuss AI and possible applications to science and data structures
  - Charles Vardeman to present AI concepts to DSC at the Nov 13th meeting.

AI – Moving Forward

- Looking for science and/or technical partners that are willing to work with OOI CI on a proof-of-concept project with a specific question to answer utilizing AI concepts.
  - Or, should a proof of concept try to recreate another AI project with OOI data (wave height prediction)?
- Should OOI develop AI models for user consumption?

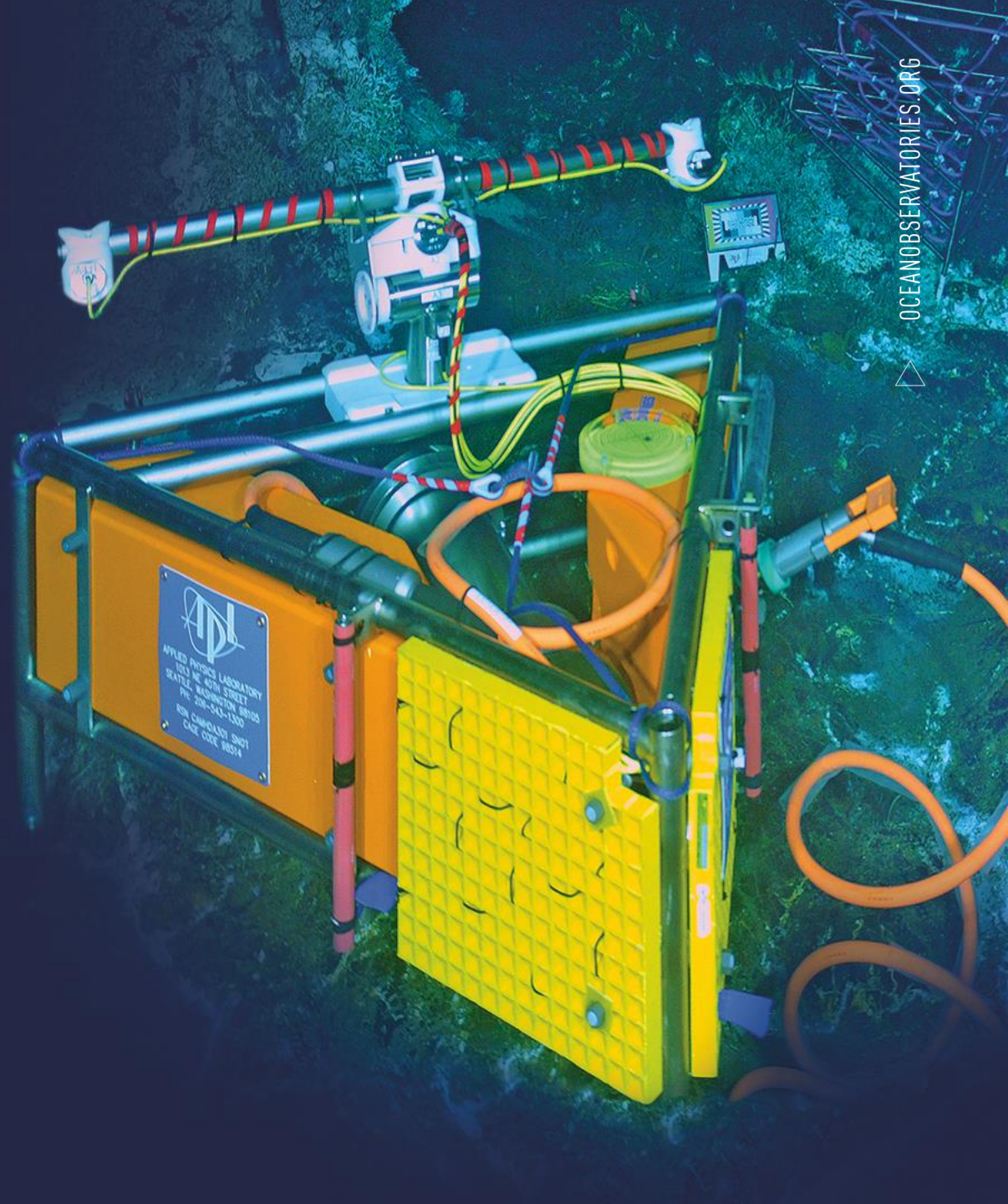


DOI Progress and Next Steps

- Current state
  - DOIs have been registered at the Site level. Site level was chosen to reduce the number of DOIs required to cite in a research paper.
  - DataCite is the service being used with all DOIs now set to ‘findable’.
- Next steps
  - Update oceanobservatories.org Site pages with citation text, link to how to site page and link to Data Explorer with Site as filter.
- Future enhancement
  - Update Stream Engine (SE) to include DOI in data files
  - Create concept of Persistent ID (PID) and add to SE for inclusion in data files.
  - Create webpage on oceanobservatories.org to lookup files associated with the Site DOI inclusive the PID. Use of both IDs will direct users to exact datasets. Effective dating will also be available.

OCEAN OBSERVATORIES INITIATIVE

Questions?



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