OCEAN OBSERVATORIES INITIATIVE

OOI Data Delivery Systems: Present and Future

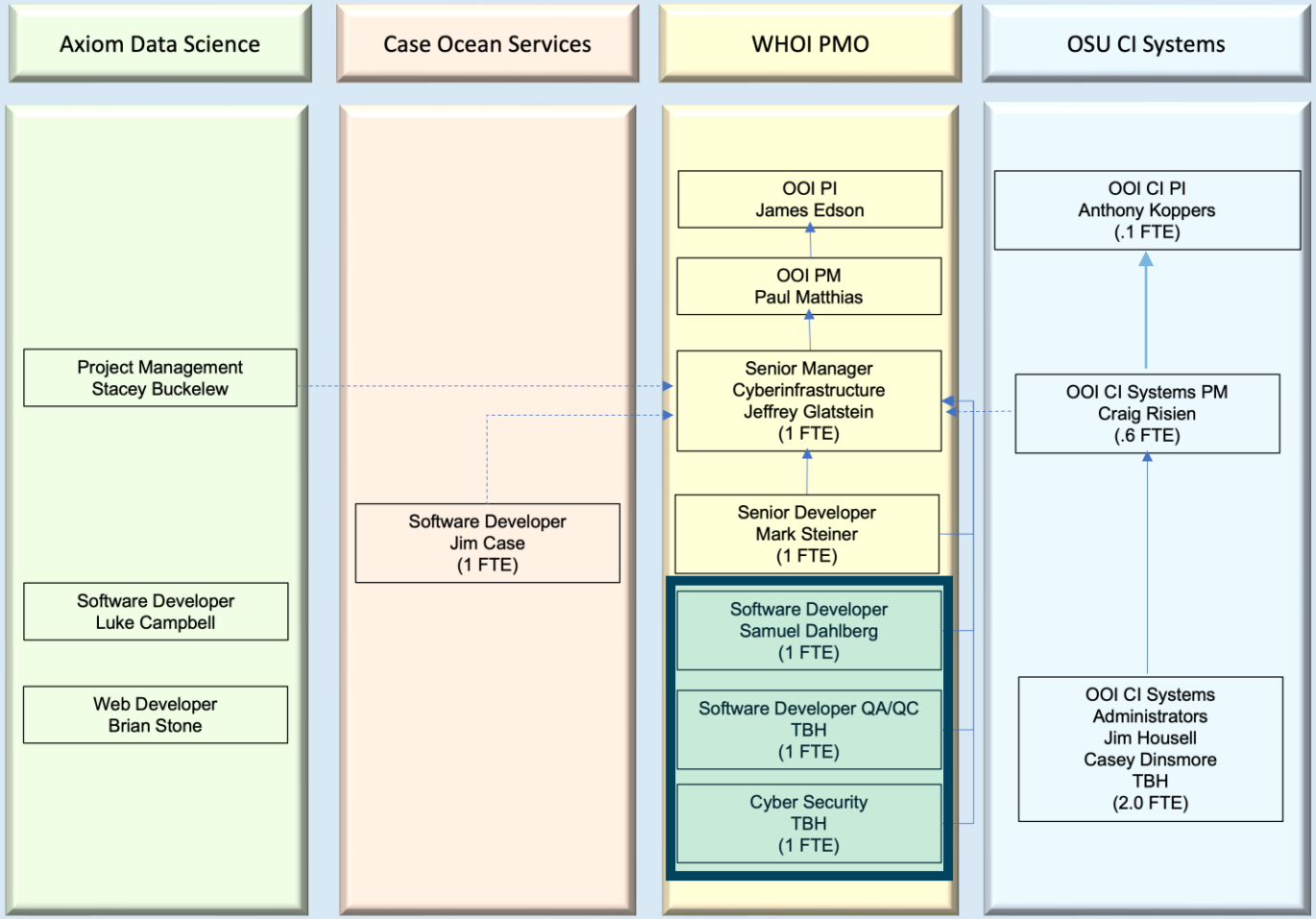
Wednesday, October 11th, 2023

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure

OCEANOBSERVATORIES.ORG

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.



Note: Software Administrator retired PYV Q2
2.5 will add CISO and Senior Software Engineer

Software Stack– High Level

- Data Processing
  - Databases (Cassandra and PostgreSQL)
  - Edex (NOAA AWIPSII branch)
  - Data Ingestion
    - Ingest engine
    - Data parsers
    - Queues by delivery method (cabled, telemetered, recovered and playback)
  - Data Delivery
    - StreamEngine/M2M
      - Preload database
      - ION functions
    - QA code (e.g. QARTOD)
- Data Discovery
  - Data Explorer
  - ERDDAP
  - Data Portal (OOINET)
  - THREDDS ‘Gold Server’
  - Raw Data Server
  - JupyterHub

Significant Projects and Impacts to Date

- Performance
  - Cassandra database tuning and cluster size increase
  - New architecture – virtualization of uFrame (part of data center move)
  - OOI software and components upgrade – edex, 3rd party software and databases
  - Implementation of StreamEngine query governor
  - Removal of worst-case scenario data retrieval as default on OOINET
  - Maintain THREDDS server of precalculated datasets
- Data accuracy and FAIR
  - Implementation of QARTOD data quality code – Gross range and Climatology
  - Data Maintenance – ability to purge and replay data by time range
  - Asset management data review
  - Preload database corrections for CF compliance
  - StreamEngine aggregation tuning
  - Resolution of data quality tickets
  - DOI strategy formulation

Significant Projects and Impacts to Date

- User Experience
  - Adjustment of OOINET interface utilizing user feedback
  - Implementation of Data Explorer with user driven use cases
  - Move to precalculated data sets (calculate on demand still available)
  - Established user feedback loops (e.g. Discourse)
  - Compute in place architecture (JupyterHub)
- Efficiency and Effectiveness
  - Data back-ups – tape, cloud, TACC, and built-in redundancy
  - NCEI long-term archiving
  - Improved Cybersecurity with Trusted CI relationship and system vulnerability scanning
  - Monitoring – effectively tells CI about issues prior to the user
  - Advanced communications plans and fostered environment of cooperation

Objectives for PYVI

- Python upgrade from 2.7 to 3.11. (multi-year)
  - Stream Engine re-architecture
    - 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
  - Parsers, ION functions and Port Agents (MI Instrument)
- Data Explorer
  - Completion of full resolution data visualization and start on three dimensional
  - Expansion of media server to include HD video, Hydrophone
  - Data Explorer operational training to OOI development and operational resources
  - Operational and performance tuning
  - Integrate JupyterHub
  - Determine plan for engineering data
- Asset management – Roundabout development
- Broader cybersecurity posture

Objectives for PYVI

- Data Accuracy and FAIR
  - Continue to target data quality tickets
  - Continue QC and QARTOD support and development of tests and tools
  - Continue to support preload database analysis and adjustments
  - Continue FAIR data standards tuning (JupyterHUB, Preload database adjustments)
- Performance
  - Query performance analysis
- Operational
  - Data Center technical refresh
  - Data Center virtualization
  - Productionalize Cloud storage transfer to TACC
  - Research NCEI data archival of raw data
  - Dev-ops, Monitoring and improved efficiency of releases
  - Database replication
  - Disaster recovery scenario exercises

Objectives for PYVI

- Strategic
  - ERDDAP tuning and replacement evaluation
  - Deliver Digital Object Identifiers (DOI) implementation (multi-year)
  - Evaluate options to reduce or eliminate the Cassandra and PostgreSQL database footprint
  - Continued cloud analysis
  - Machine learning and AI research and proof of concepts

Potential DSC Topics

- 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?