Cyberinfrastructure (CI) Systems PYIII Accomplishments and PYIV Planned Activities

Anthony Koppers (PI), Craig Risien (PM)

Jim Housell (IT Architect), Casey Dinsmore (DevOps)

Oregon State University

OCEAN OBSERVATORIES INITIATIVE

OCEANOBSERVATORIES.ORG

OSU CI Systems Proposal Submitted June 2020

*"Our overarching goals are to create a **low-risk transition** from the OOI-CI operations at Rutgers University to OSU in Year 1, **without downtime** but with significantly **increased capabilities and security**, while setting the stage for **significant extensibility** in OOI-CI service during Years 2 & 3 — and beyond."*

Toward an Extensible Cyberinfrastructure for the Ocean Observatories Initiative

Lead-PI: Anthony A.P. Koppers, Professor and ADRO, CEOAS, Oregon State University
Co-PI: Mark Keever, RCS Director, CEOAS, Oregon State University
OOI-CI Operations Manager: Craig Risien, OOI, CEOAS, Oregon State University

1. INTRODUCTION

The Ocean Observatories Initiative (OOI) serves the larger Ocean Sciences community in the U.S. and the world. Its research themes cover atmosphere-ocean interactions, climate change, ocean mixing and life in the oceans, coastal systems and ecosystems, interactions between ocean water and the subseafloor, microbial life in the ocean realm, and large-scale plate tectonic, earthquake and volcanic processes. OOI collects data from the ocean floor to the atmosphere using infrastructure that includes surface and subsurface moorings, cabled seafloor platforms, profilers, gliders, and autonomous underwater vehicles. These marine-based platforms deliver real time data to shore via cable, satellite, and cell phone technologies. OOI instruments range from well-tested commercially-available sensors to unique sensors added by external Principal Investigators. OOI instruments have widely varying formats, bandwidth requirements, quality control requirements, and metadata.

While the instrumentation in the five research arrays forms OOI's backbone, the data collected, processed, stored, archived, and made available for public access, are its heart. The OOI Cyberinfrastructure (OOI-CI) provides real time access to data collected by OOI from its inception to date. With every year, the cumulative data sets become more meaningful. As researchers get the opportunity to work with longer and more complex time series, synthesis of these data will be key. The OOI-CI needs to be dependable, always on, fast, and streamlined for future research and unanticipated needs in data analytics. The data also needs to be archived in accord with NSF data policies and to be available throughout OOI's lifetime and after its sunset in the distant future. Finally, the OOI-CI should be extensible to seamlessly cope with the expected growth and increases in sampling frequency in support of future OOI science projects.

OCEAN OBSERVATORIES INITIATIVE

NSF

Proposed OSU CI Systems Architecture
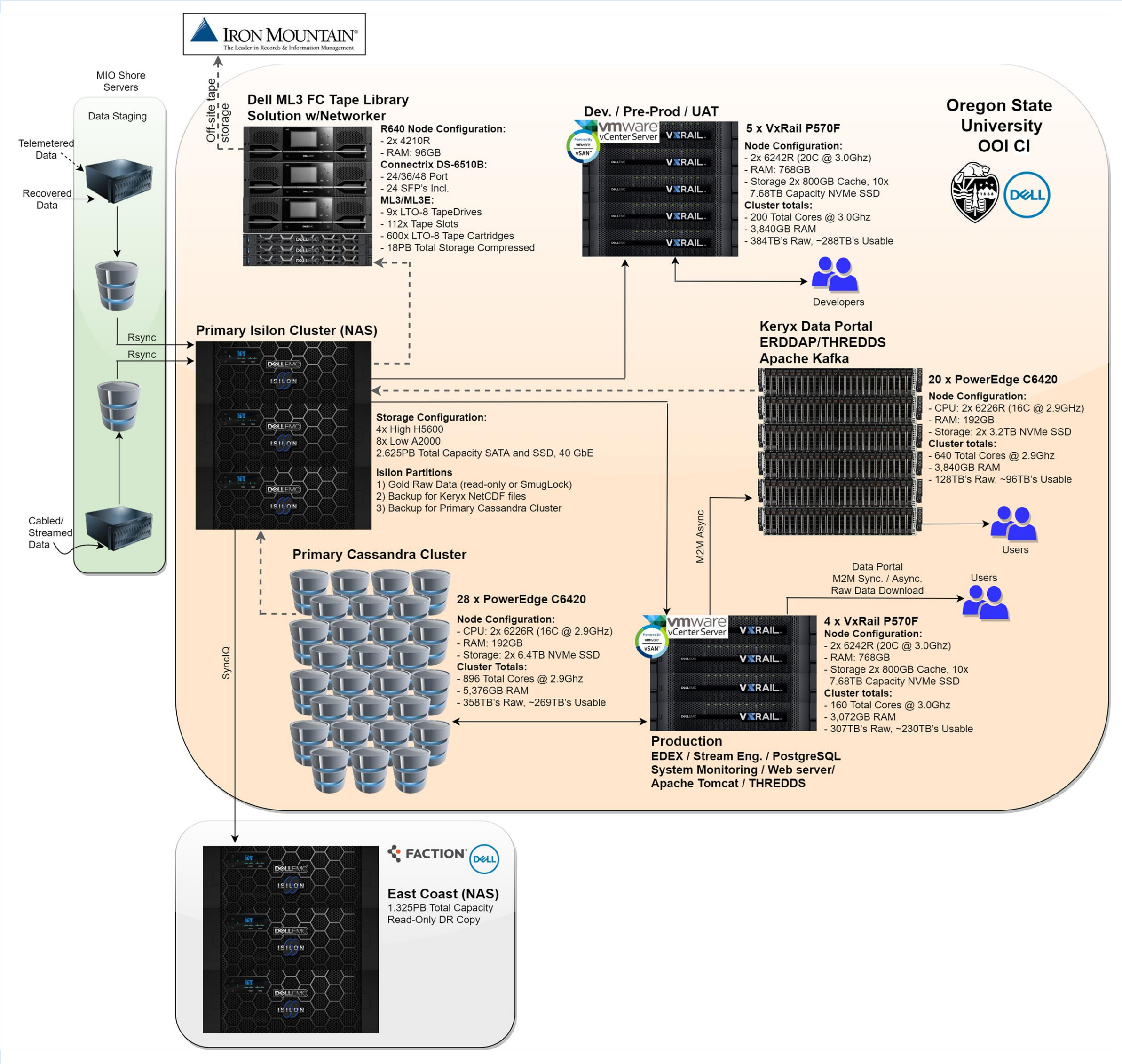
Cyber Security and Data Protection

- Palo Alto Firewall

- Single VPN for OOInet

- Duo 2-factor Authentication

Connectivity

- Independent Dual Routing, 100 Gb

- LinkOregon connections to Internet2

OSU Data Center

- Redundant PDUs and cooling

- Backed up by diesel generator

- Access control, fire monitoring

Proposed OSU CI Systems Architecture

ON-PREM COMPONENTS	STORAGE	CPU	RAM	FEATURES	PROPOSED USE
Primary Isilon Cluster (NAS) Dell EMC H5600 + A2000 Copy 1 :: West Coast (Corvallis)	2,625 TB			• SATA / SSD Storage • 10-40 GbE • 3x Build-in Replication natively in Isilon	• Read-only Raw Data Scale-out Storage • On-prem backup for Keryx NetCDF Data Farm and Cassandra Cluster • Total storage (100%) over 3 years
Primary Cassandra Cluster PowerEdge C6420, 28 nodes	358 TB	• 896 cores • 2.9 GHz • 6226R	5,376 GB	• NVMe SSD Storage • 3x Build-in Replication in Cassandra Software	• Installed over two OSU on-campus cold rooms (one in ECC; another in OSU-IT)
Keryx Kafka Data Portal Cluster PowerEdge C6420, 20 nodes	128 TB	• 640 cores • 2.9 GHz • 6226R	3,840 GB	• NVMe SSD Storage	• ERDDAP, THREDDS • NetCDF File Farm
Production VM Server VxRail P570F, 4 nodes	307 TB	• 160 cores • 3.0 GHz • 6242R	3,072 GB	• NVMe SSD Storage	• Run all compute processes as VMware in vSphere/vSAN • EDEX, Stream Engine, PostgreSQL, System Monitoring, Web Server, Tomcat, THREDDS, etc.
Dev/Test VM Server VxRail P570F, 5 nodes	384 TB	• 200 cores • 3.0 GHz • 6242R	3,840 GB	• NVMe SSD Storage	• Run all dev/test processes as VMware in vSphere/vSAN • Allows identical VMs for Development, Pre-production, UAT

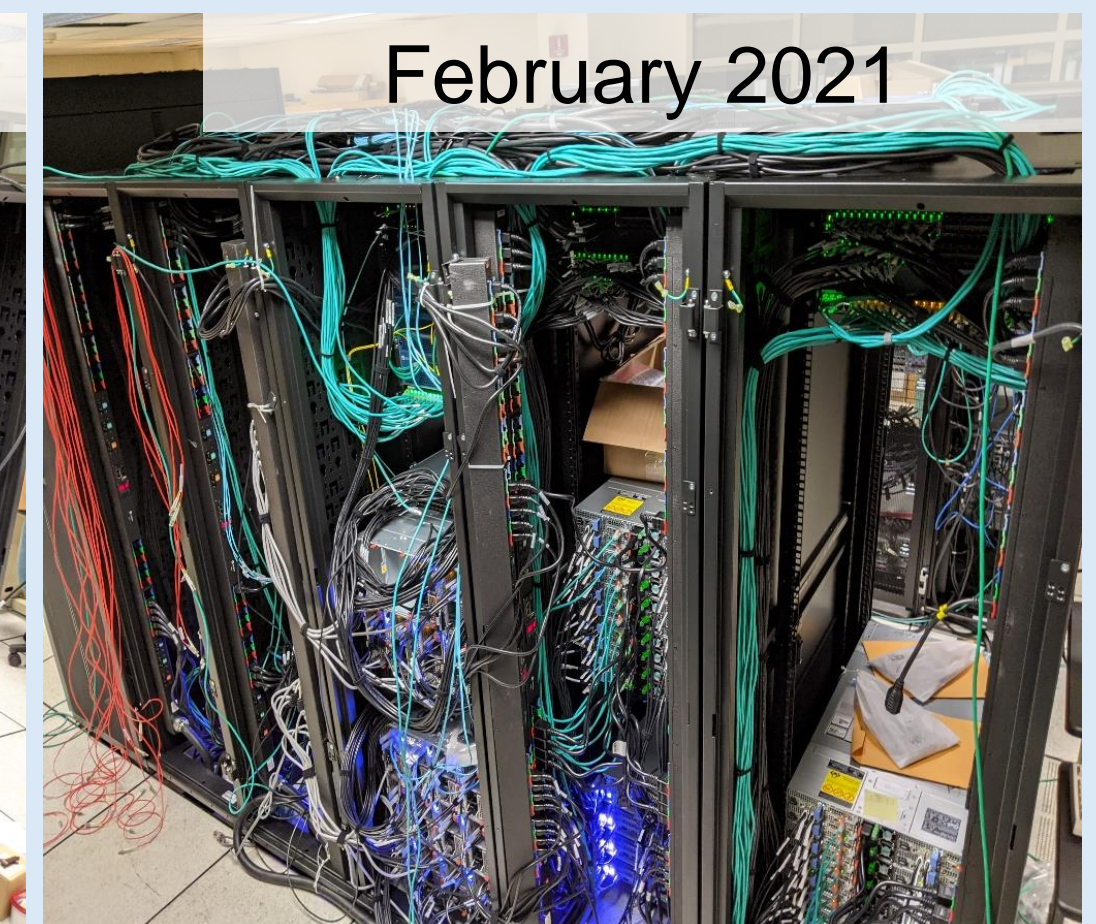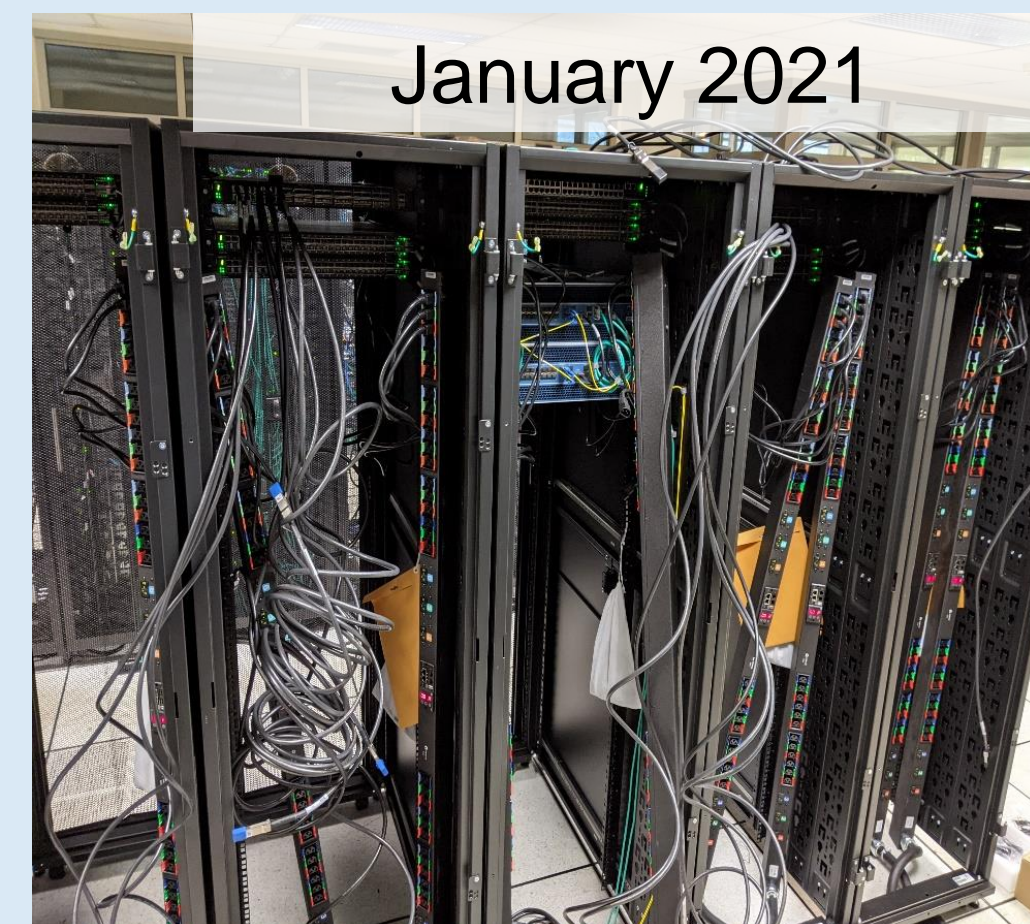| BACK-UP COMPONENTS	STORAGE	CPU	RAM	FEATURES	PROPOSED USE
Secondary Isilon Cluster in Cloud hosted by Faction Dell EMC H5600 + A2000 Copy 2 :: East Coast	1,325 TB			• active:active replication on Isilon Cluster in East Coast Faction Data Center	• Could be made primary during DR • Total storage (100%) over 3 years
Dell ML3 FC Tape Backup and Offsite Iron Mountain Storage Copy 3 :: West Coast (Salem)	18,000 TB	• 20 cores • 2.4 GHz • 4210R	96 GB	• Networker • Fiber Cable • 12 Sockets	• Daily backup on tape of *all storage* on the Primary Isilon NAS Cluster • Weekly shipment to Iron Mountain

Data Storage

- Primary Isilon cluster NAS 3d:1n1d Razor Protection

- 28 node Cassandra and 20 node Keryx clusters

- VxRail servers w/ VmWare

Disaster Recovery

- Secondary Isilon East Coast

- Tape backup with one onside and two offside copies

OCEAN OBSERVATORIES INITIATIVE

OSU CI Systems PYIII Accomplishments

- **Oct 2020:** MON approved; Hardware POs submitted to Dell
- **Nov 2020:** Hardware delivered; Installation planning began
- **Dec 2021:** Partial (450 TB) copy of raw data arrives at OSU; Installation planning completed
- **Jan-Feb 2021:** All Dell **Hardware Installed**
- **Mar-Apr 2021:** Production VMs created & software installed; Cabled data migration from CI-E to CI-W initiated
- **May 2021:** Copying of **Raw Data Completed**; Cabled data migration completed
- **Jun 2021:** Uncabled data ingestions completed; Faction Isilon moved to east coast data center
- **Jul 2021:** OSU became **OOI System of Record on July 30**
- **Aug 2021:** Data Explorer live on OSU hardware
- **Sept 2021:** Rutgers's decommissioning completed

October 2020

November 2020

January 2021

February 2021

OCEANOBSERVATORIES.ORG

OOI OCEAN OBSERVATORIES INITIATIVE

NSF

OSU CI Systems PYIV Planned Activities

Cyber Security and Data Protection

- Implementing Trusted CI framework

- Application of CIS baseline control set

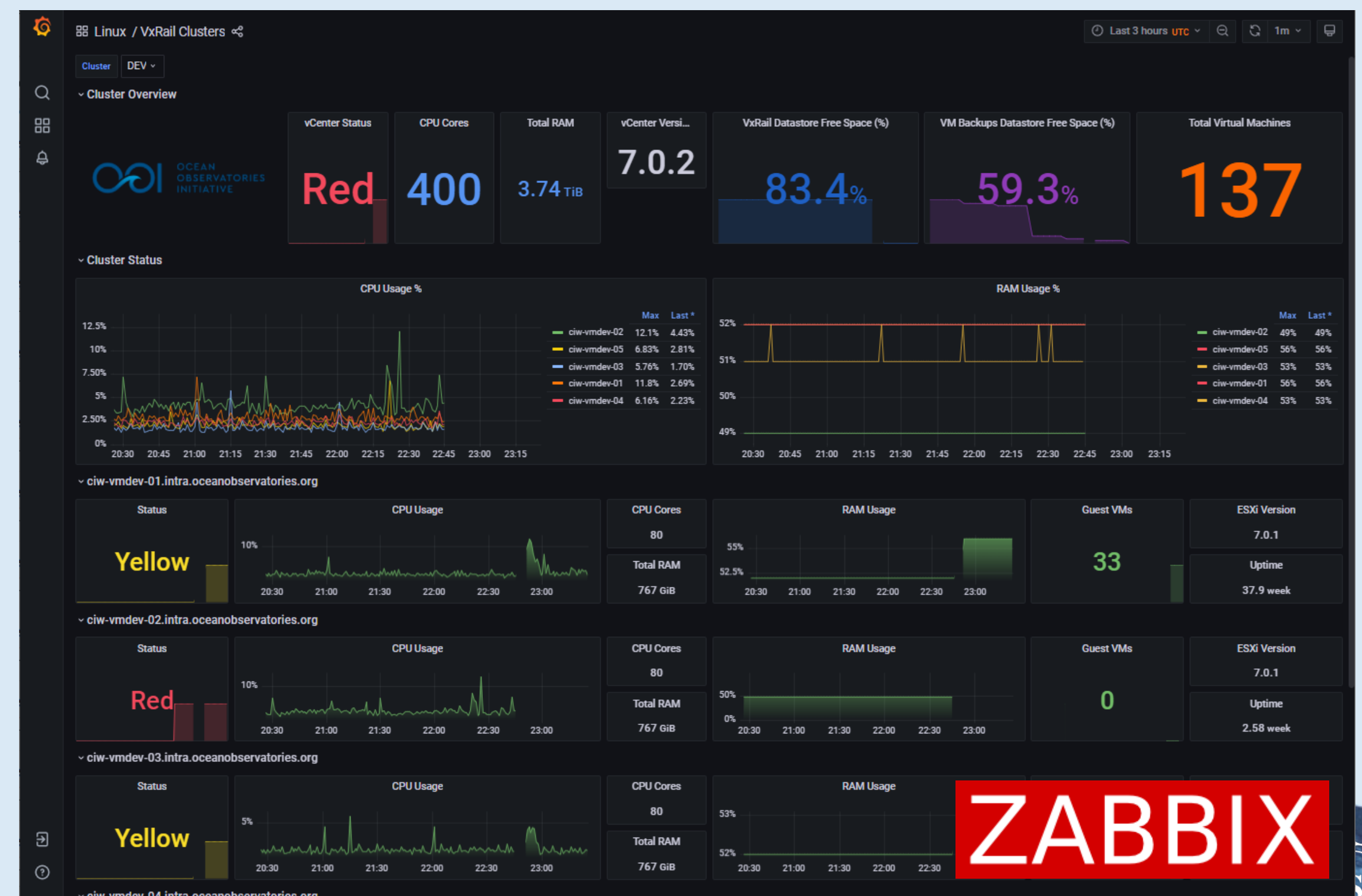- Vulnerability scanning using Tenable

System Monitoring and Alerts

- InsightIQ and DataIQ

- Zabbix agents

- Grafana dashboards

Cloud computing POC

NOAA-NCEI deep data retention agreement
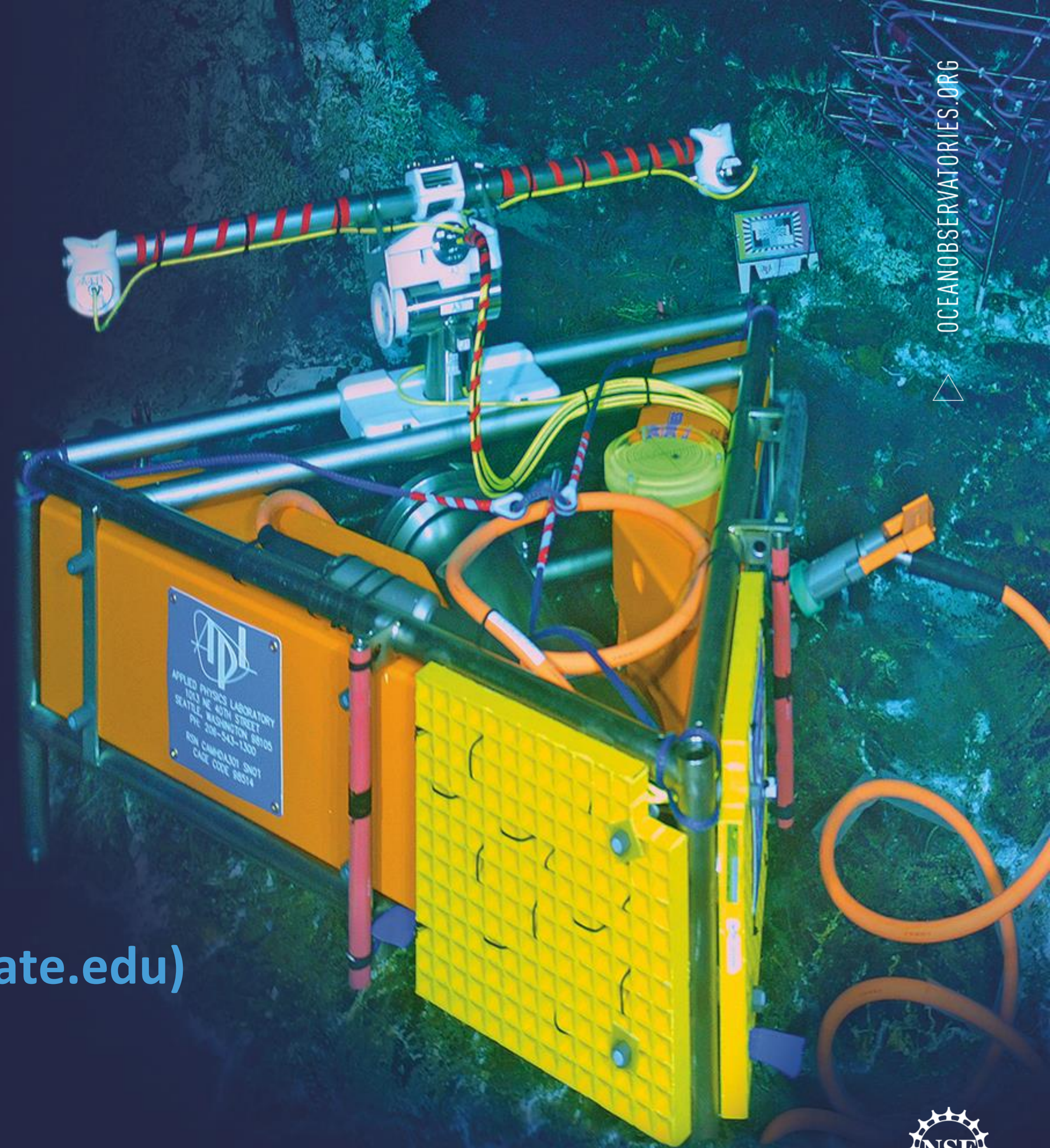
System maintenance and support

Questions?

Anthony Koppers (anthony.koppers@oregonstate.edu)

Craig Risien (craig.risien@oregonstate.edu)

OCEAN
OBSERVATORIES
INITIATIVE

OCEANOBSERVATORIES.ORG