



Cyberinfrastructure Update

Jeffrey Glatstein

Senior Manager of Cyberinfrastructure

Anthony Koppers

Principal Investigator OSU Data Center

Northeast Pacific Workshop

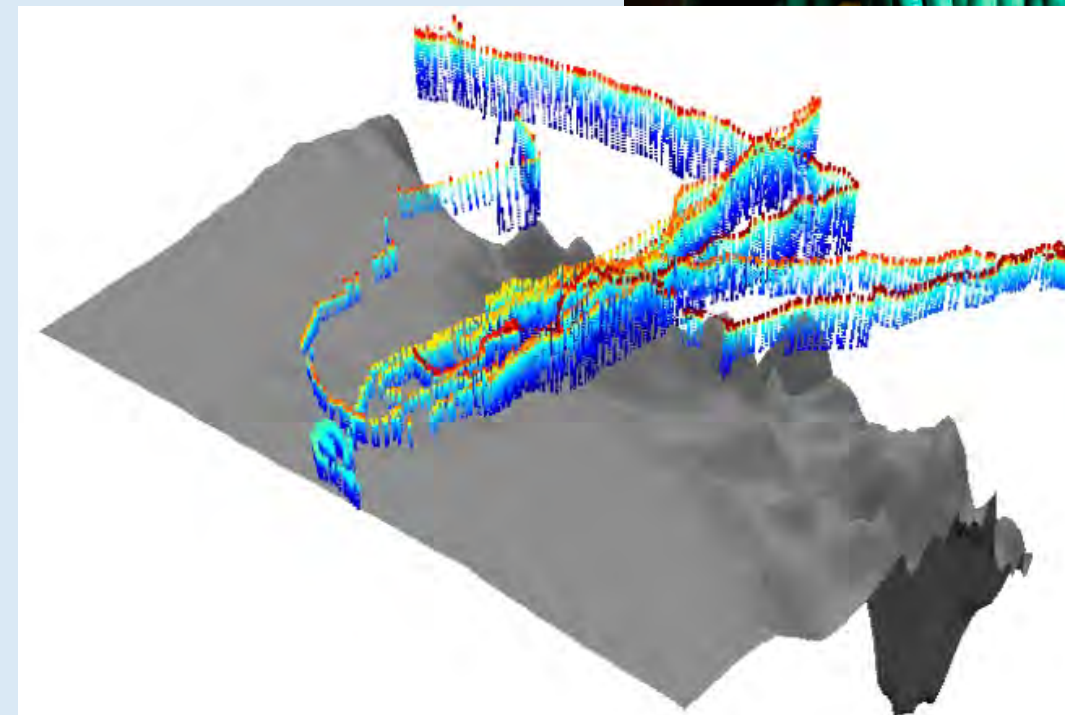
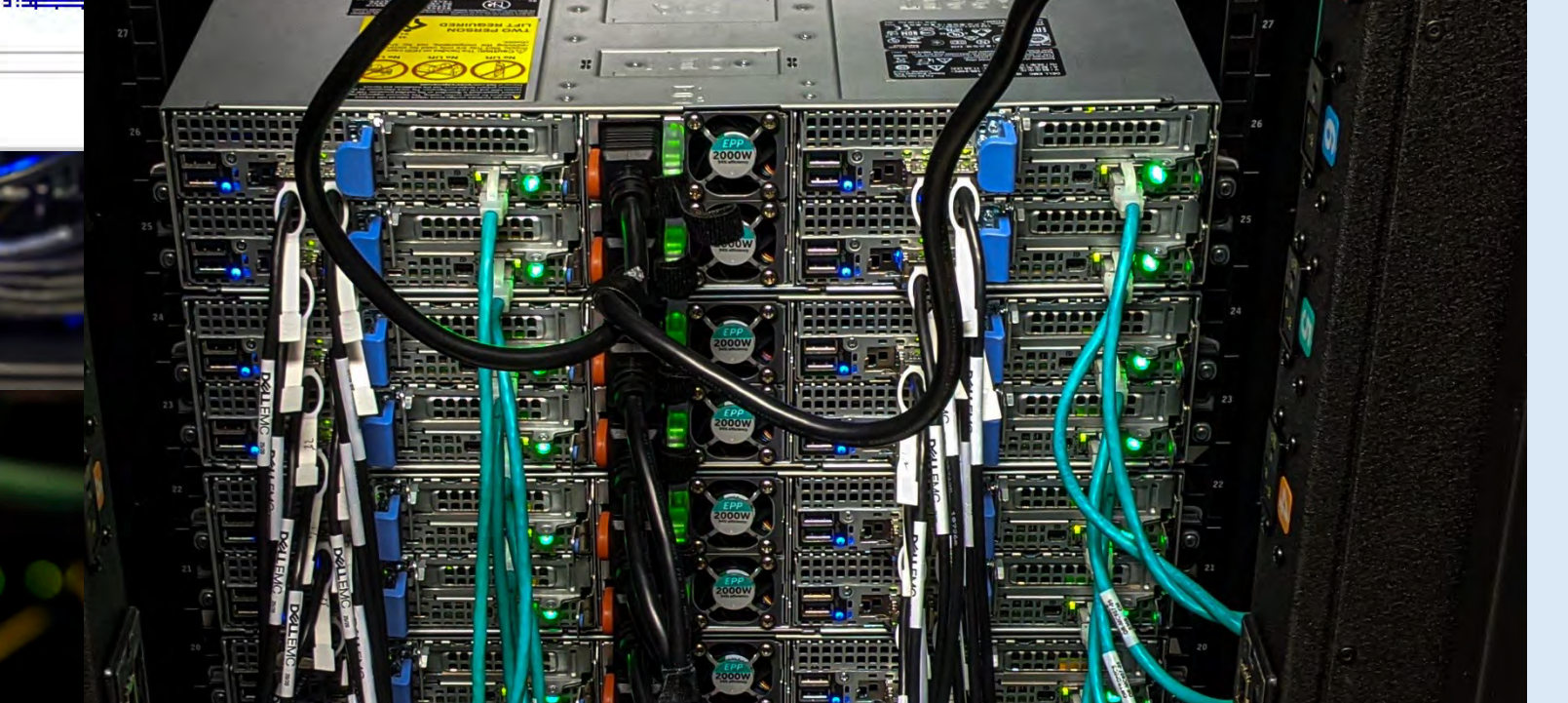
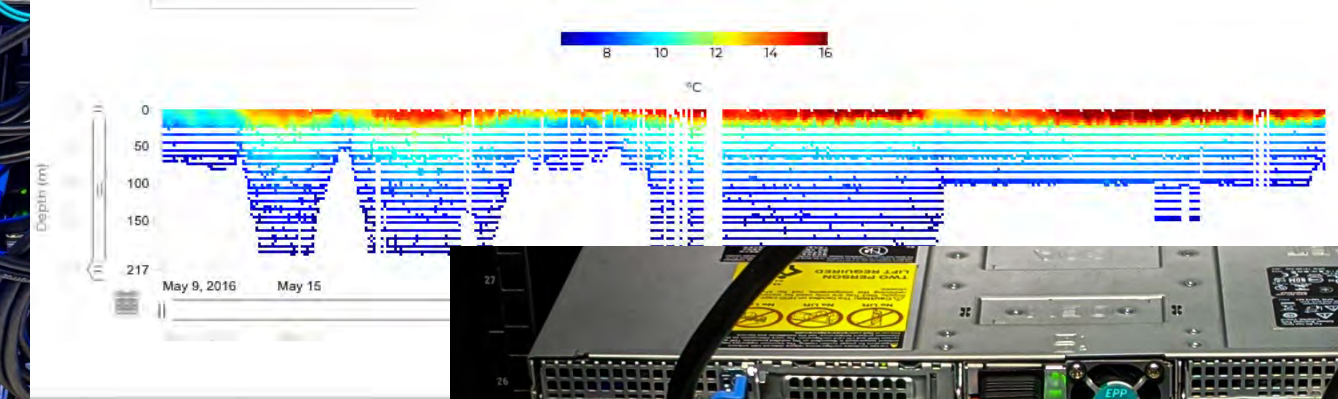
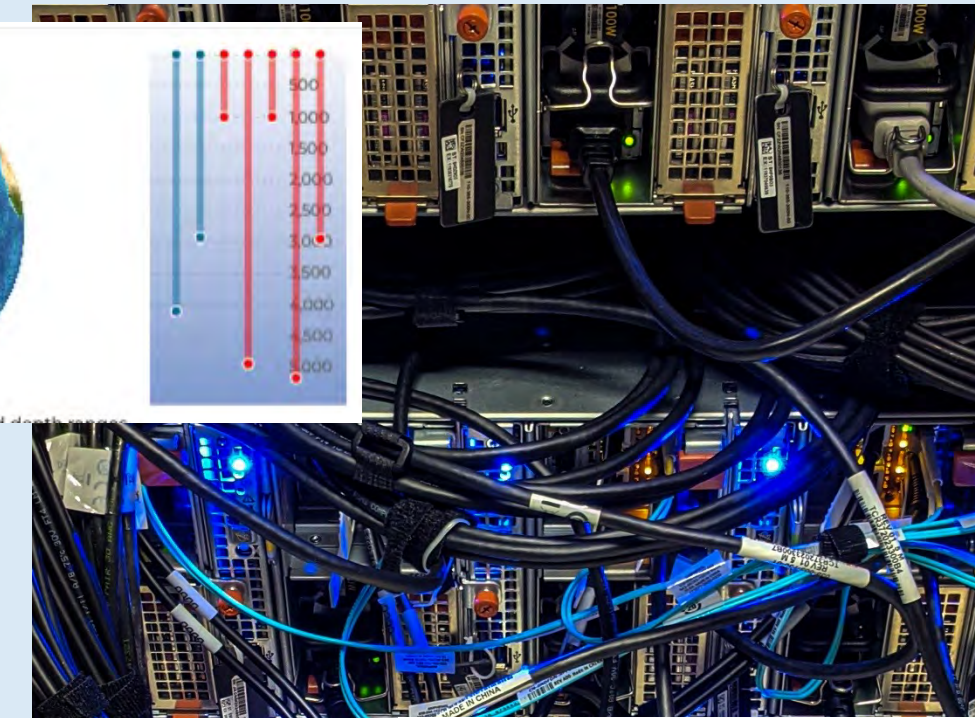
June 7, 2022



Agenda

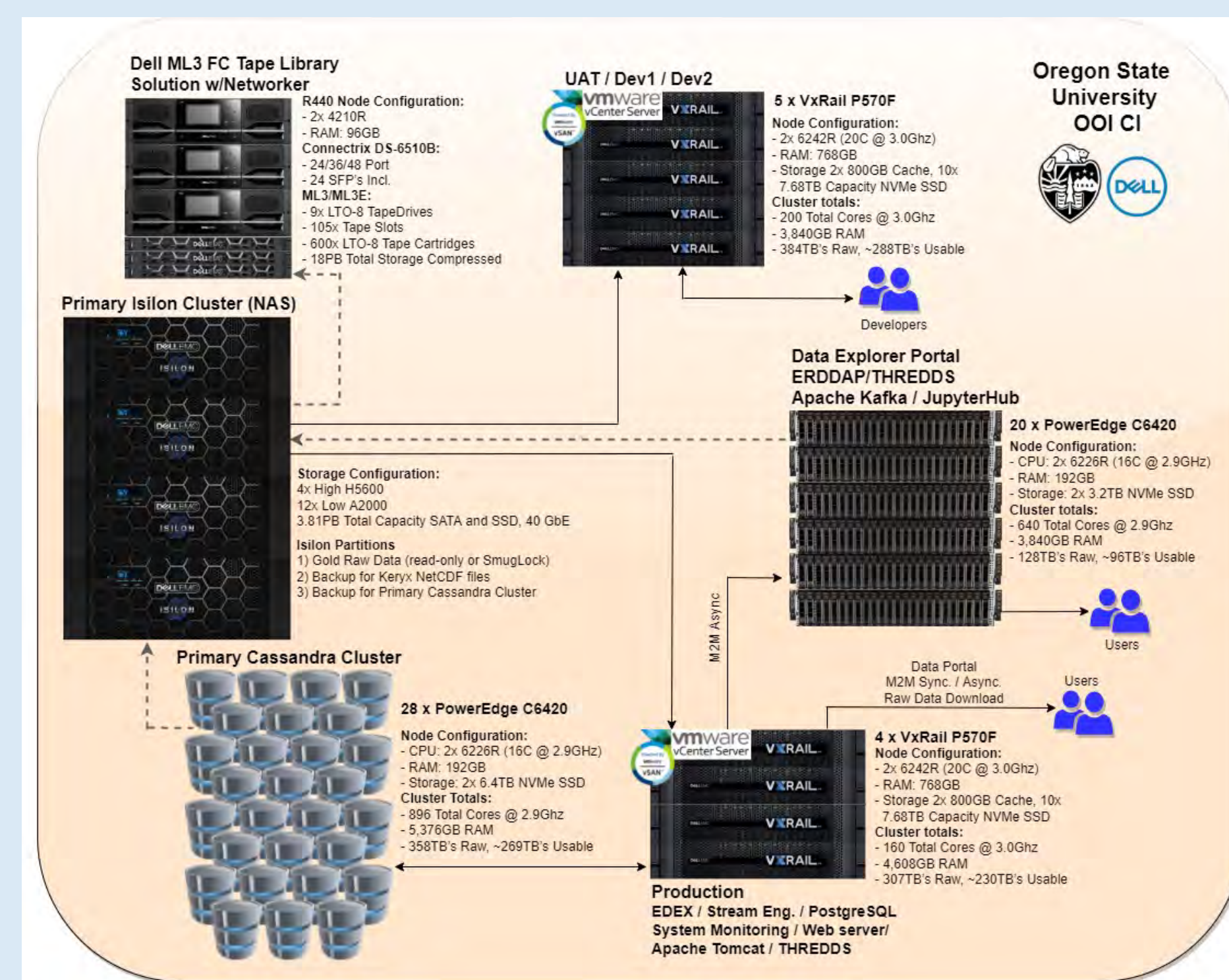
Discuss what is new in the cyberinfrastructure area:

- OSU Data Center
- Cybersecurity & Data Protection
- System Monitoring
- New Data Explorer Features
- How to Get Help
- Data Interfaces
- JupyterHub Demo



OSU Data Center

- OOI system of record since July 30, 2021
- Isilon Storage: 3.81PB
- VxRail Compute: 360 Cores; 8.4TB RAM; 684TB Storage
194 Virtual Machines (Prod, UAT, Dev1, Dev2)
- DataExplorer / JupyterHub: 640 Cores; 3.8TB RAM; 128TB Storage
- Prod. Cassandra Cluster: 896 Cores; 5.4TB RAM; 360TB Storage
- 2 x Palo Alto Networks Next-Gen Firewalls
- 14 x Dell 100GbE PowerSwitches



Cybersecurity

- Working with Trusted CI (TCI) to adopt and implement the TCI Cybersecurity Framework.
- Duo Multi-Factor Authentication (MFA) for Virtual Private Network (VPN) connections to OOI-CI Palo Alto firewalls.
- Implementing Center for Internet Security (CIS) Critical Security Controls.
- Internal and external vulnerability scanning using Tenable.io

Data Protection

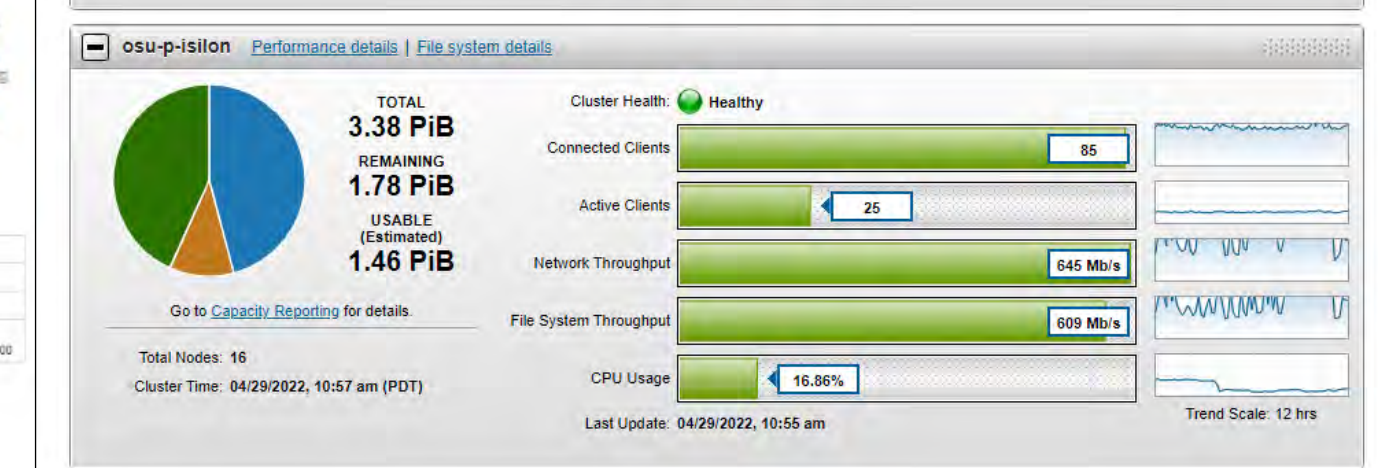
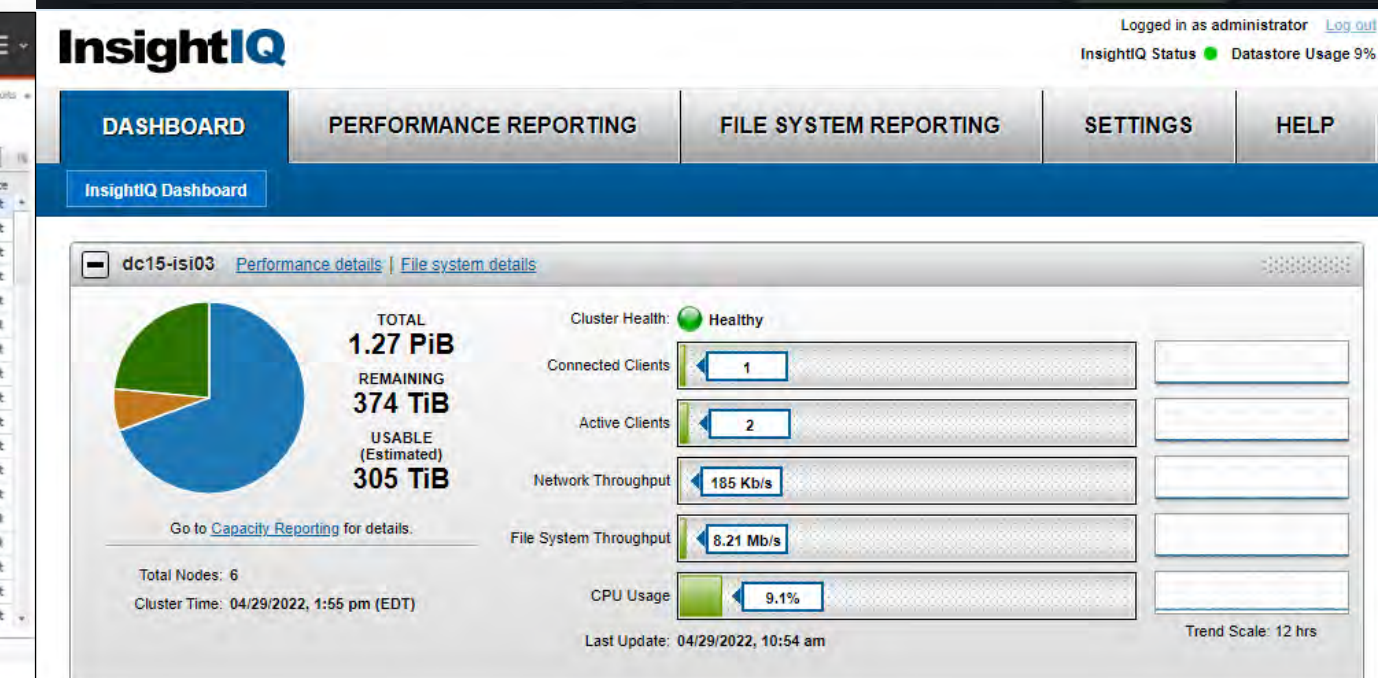
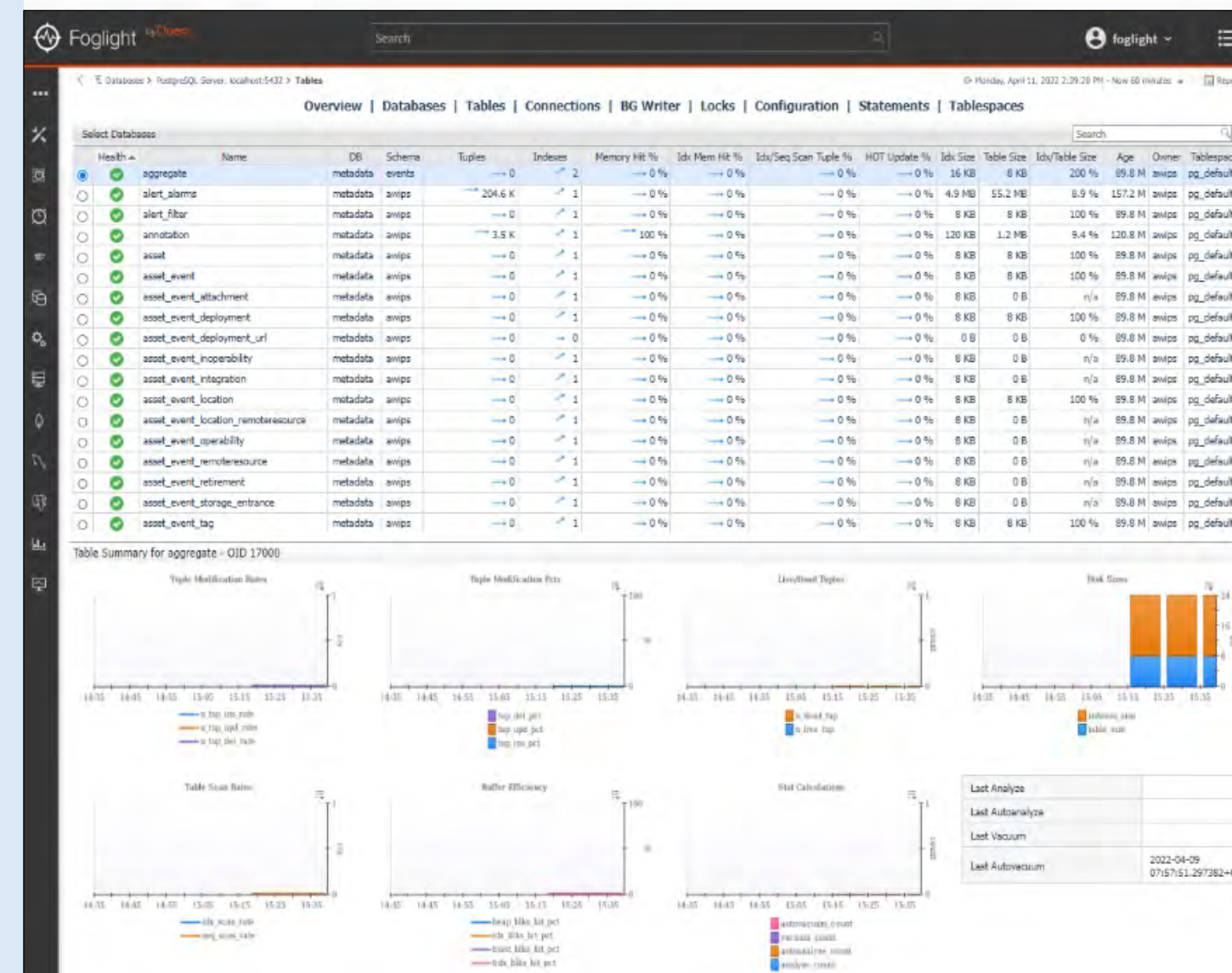
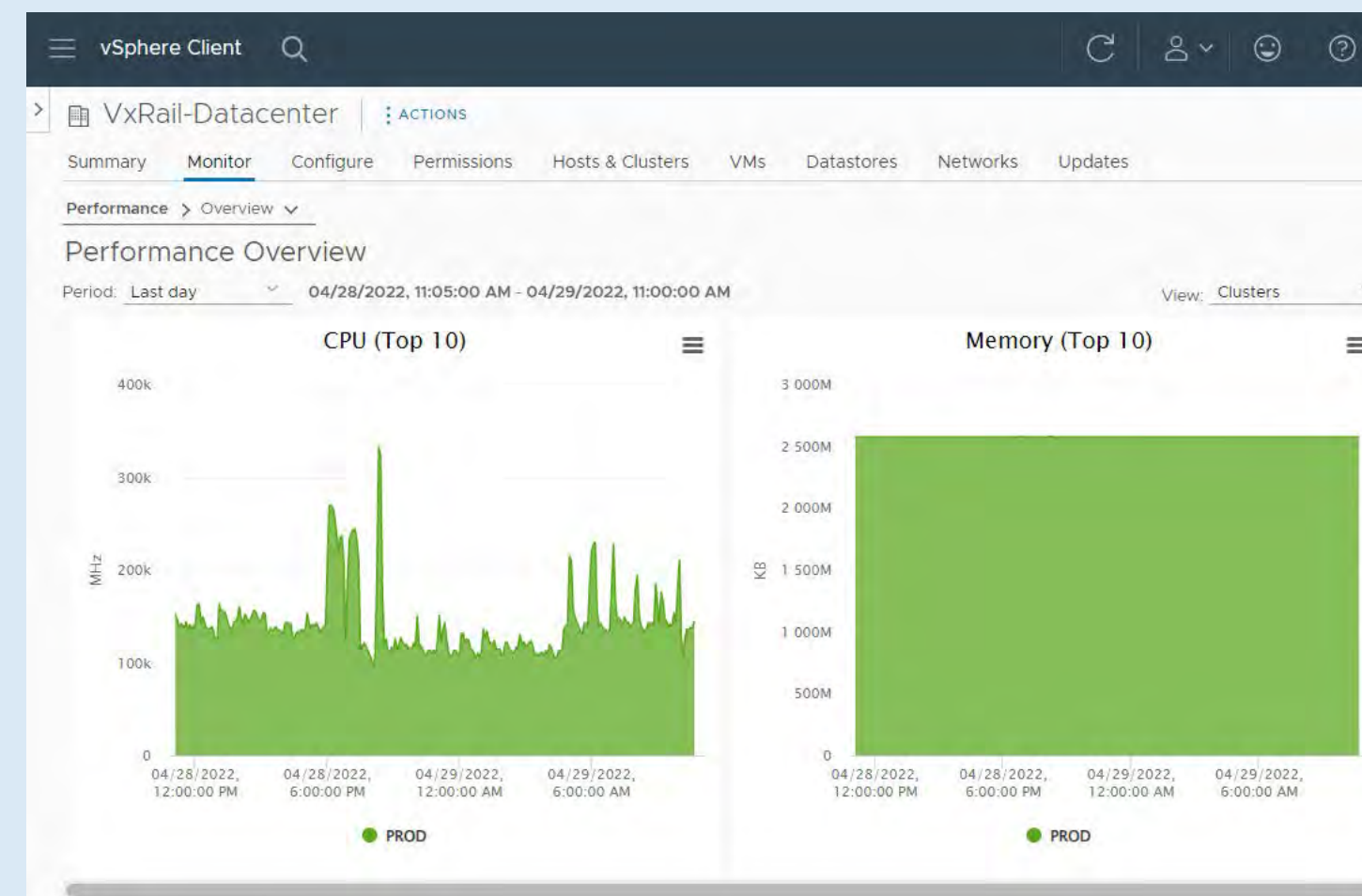
- Isilon data protection level is set to +3d:1n1d, meaning that the cluster can survive three simultaneous disk failures or one entire node failure plus one disk.
- Daily data backups to a US East Coast Faction Inc. data center and to local LTO tape.

The image shows two screenshots. The top screenshot is the 'The Trusted CI Framework' website. It features the Trusted CI logo (The NSA Cybersecurity Center of Excellence) and navigation links: Blog, Resources, Events, Success Stories, About, Contact. The main heading is 'The Trusted CI Framework'. Below it, there's a description: 'The Trusted CI Framework is a minimum standard for cybersecurity programs. In response to cybersecurity guidance focused narrowly on cybersecurity controls, the Trusted CI Framework provides a more holistic and mission-focused standard for managing cybersecurity.' To the right is a 'Cybersecurity Program' badge that says 'BUILT WITH THE TRUSTED CI FRAMEWORK' and a 'DOWNLOAD OUR BADGE' button. On the left, there's a sidebar with links: About, Framework Core, Implementation Guidance, Templates and Tools, and Share Your Feedback. Below these links is a form to 'Let us know you're using the Framework by entering your email below.' with a 'SUBMIT' button and a 'DOWNLOAD THE FRAMEWORK IMPLEMENTATION GUIDE' button. At the bottom, there are logos for NOIR Lab, GAGE, OCEAN OBSERVATORIES INITIATIVE, NRAO, NSO, FABRIC, and LIGO.

The bottom screenshot is the Tenable.io interface. The top navigation bar shows 'tenable.io' and 'Asset View > Assets > Asset Details'. The main heading is 'ciw-keryx-prod-01'. Below this, there's a table with 'Asset Information' and 'Additional Information'. The 'Asset Information' table has columns for 'IPV4 ADDRESSES', 'IPV6 ADDRESSES', 'OPERATING SYSTEM' (Linux Kernel), 'TYPE' (general-purpose), and 'DNS ENTRIES' (ciw.keryx.prod...). The 'Additional Information' table has columns for 'INSTALLED SOFTWARE' and 'TENABLE UUID'. To the right of the tables, there's a 'Vulnerability by Severity' section showing counts for CRITICAL (0), HIGH (0), MEDIUM (0), and LOW (0). Below this is a 'Tags' section with a dropdown for 'Departmental: OOI'. The bottom section is 'Scan Information' with columns for 'FIRST SEEN', 'LAST SEEN', 'LAST AUTH SCAN', and 'SOURCE'. The 'FIRST SEEN' column shows '04/01/22 at 11:19 AM'. The 'LAST SEEN' column shows '04/29/22 at 9:37 AM'. The 'LAST AUTH SCAN' column shows '04/29/22 at 9:37 AM'. The 'SOURCE' column shows 'N/A'. At the bottom, there's a table with columns for 'EVENT', 'DATE', and 'SOURCE'. The 'EVENT' column shows 'Asset Updated', 'Asset Seen', 'Asset Seen', 'Asset Seen', 'Asset Seen', and 'Asset Seen'. The 'DATE' column shows '04/29/22 at 9:37 AM', '04/28/22 at 2:38 PM', '04/27/22 at 2:34 PM', '04/26/22 at 2:37 PM', '04/25/22 at 2:35 PM', and '04/24/22 at 2:35 PM'. The 'SOURCE' column shows 'N/A' for all events.

System Monitoring

- Zabbix, Grafana, and Nagios for overall data center monitoring.
- Quest Toad Edge and Foglight for database management and monitoring.
- InsightIQ for Isilon storage performance monitoring and reporting.
- VMware vSphere for VxRail management and monitoring.
- Panorama for firewall management and monitoring.

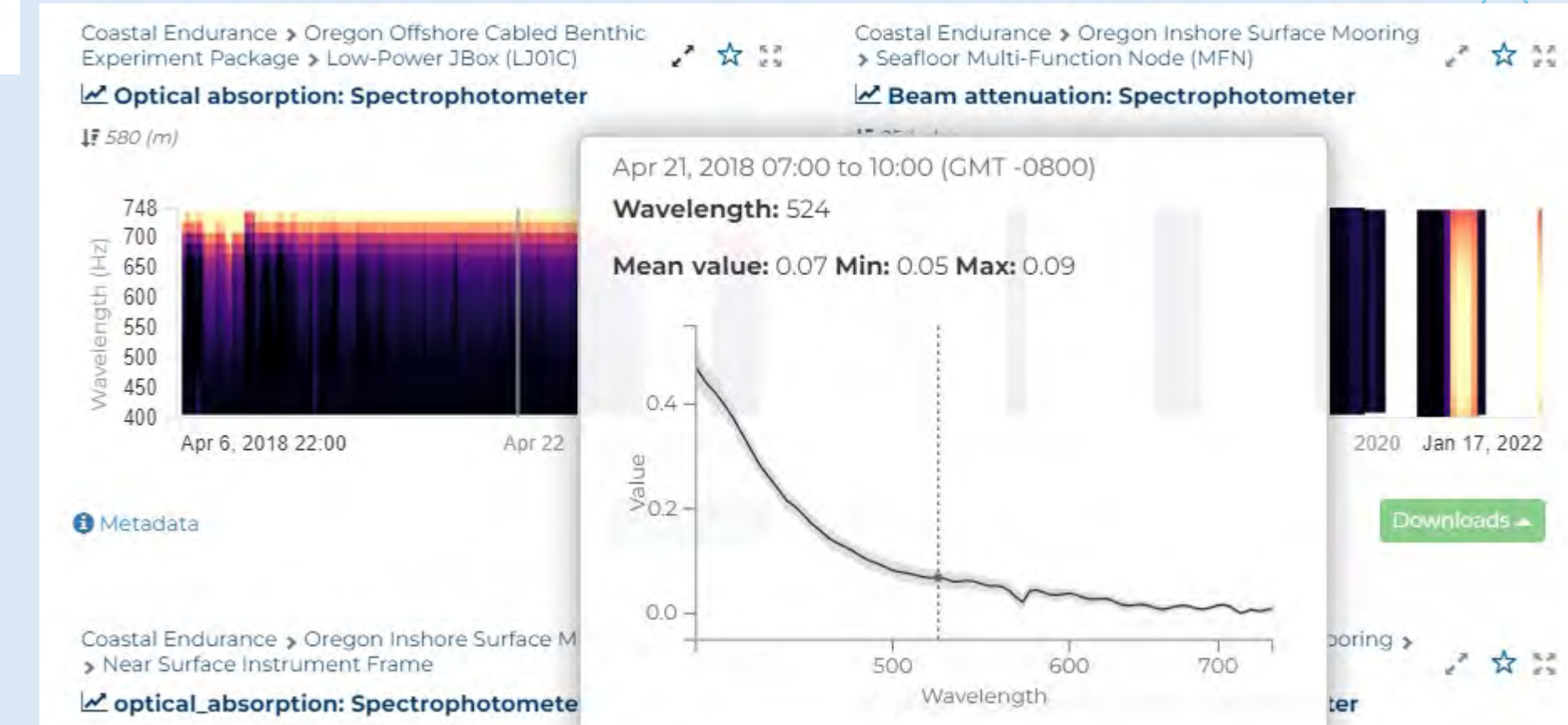
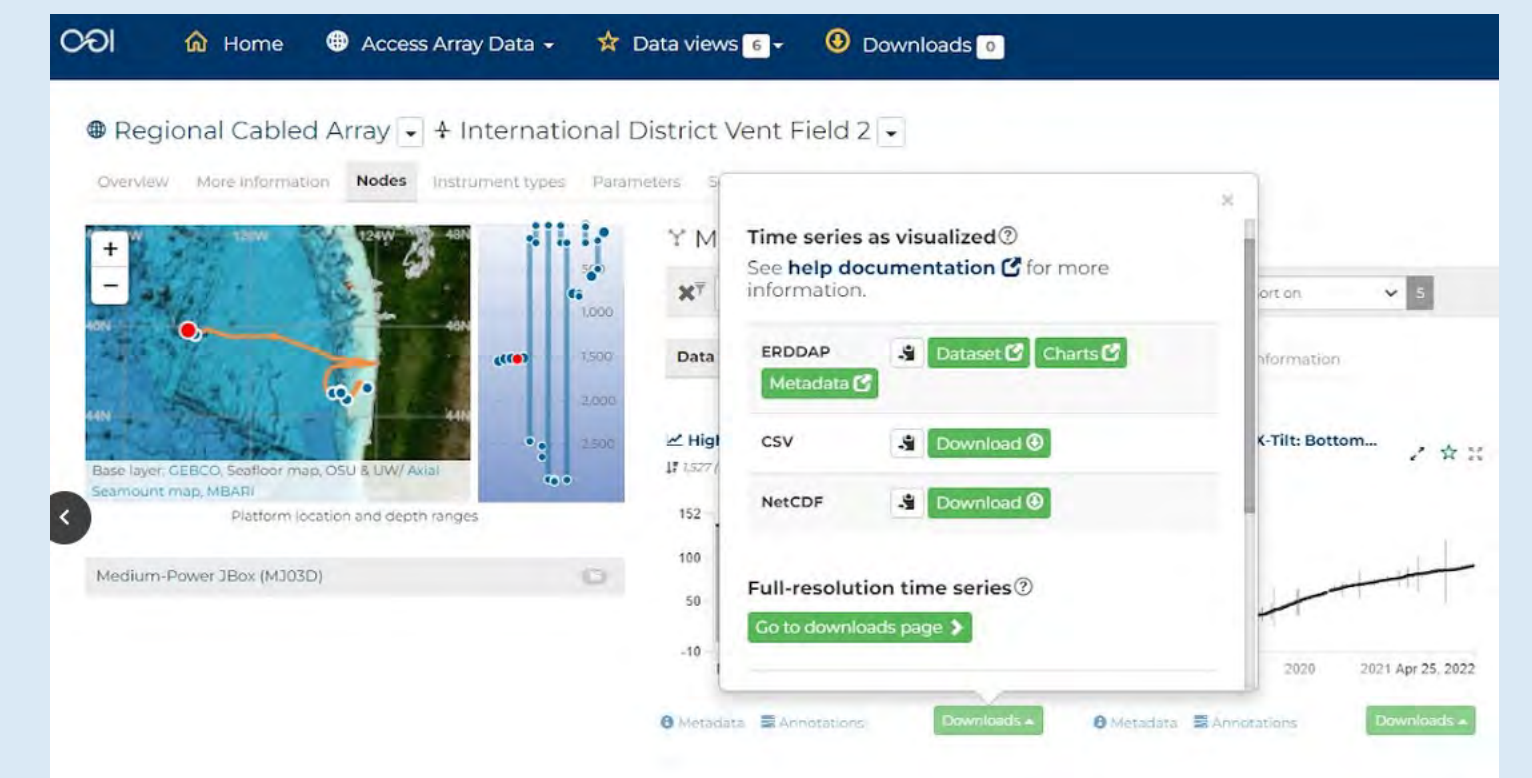
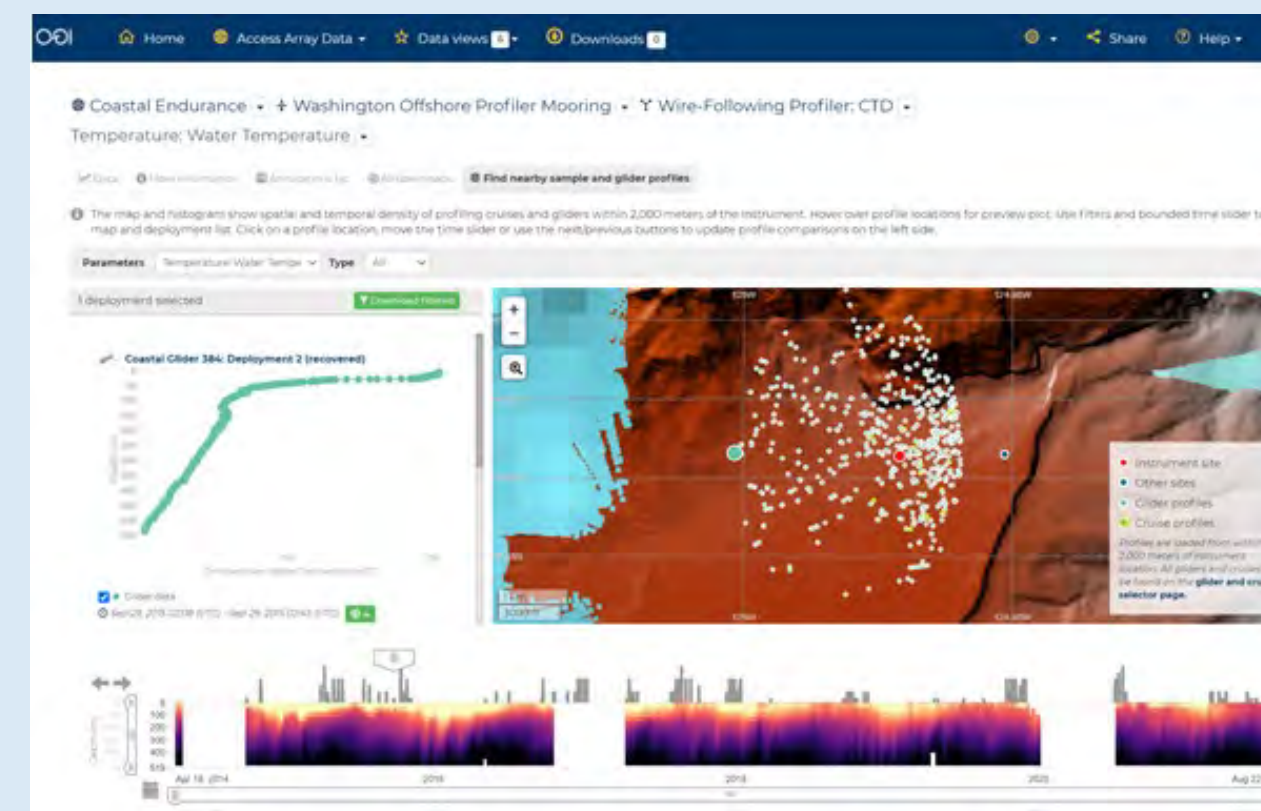
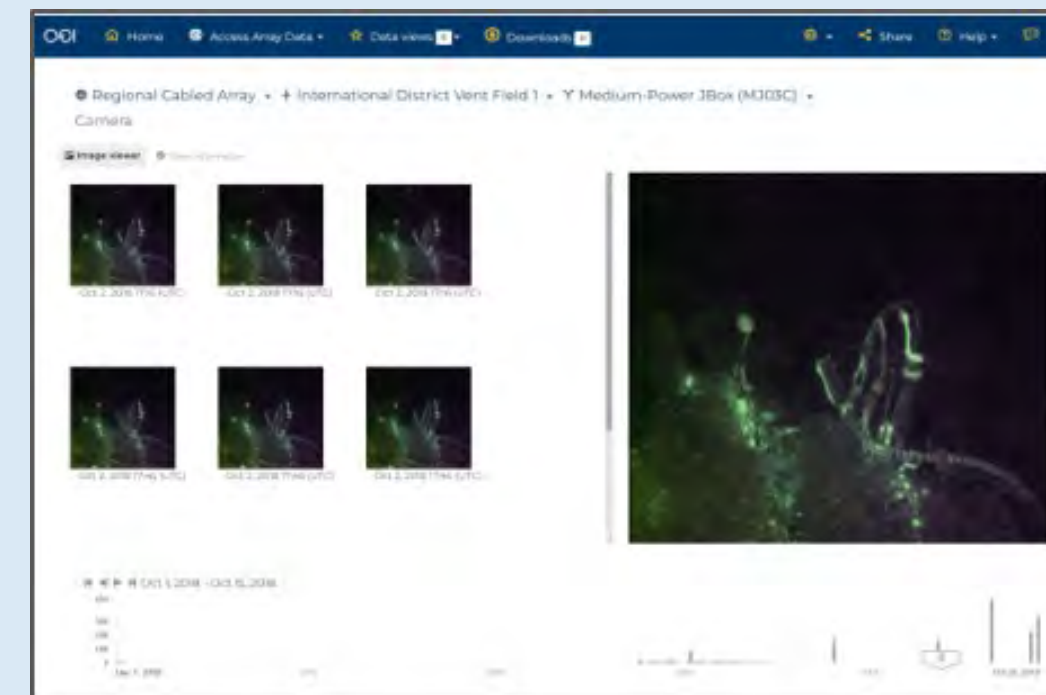


New Data Explorer Features

- Discrete Sample data available
- Glider and Discrete data search grid
- Custom polygon search capability helps visually find and compare instrument measurements to profiles collected by gliders and cruises near to instrument location
- Inset map displays data location within profile
- Clarification on data download source options
- Interactive plotting of OPTAA data
- Significant backend performance and stabilization updates

Coming Soon...

- Media server – HD pictures and videos
- Realtime data plotting
- Full resolution data visualization
- JupyterHub



```

jupyter QA_Analysis Last Checkpoint: 11/24/2021 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [44]: import requests
         from requests.auth import HTTPBasicAuth
         from timeit import default_timer as timer
         from datetime import datetime
         from datetime import timedelta
         import os
         from time import sleep
         from urllib.request import urlopen
         from urllib.request import urlopen
         from xml.dom import minidom

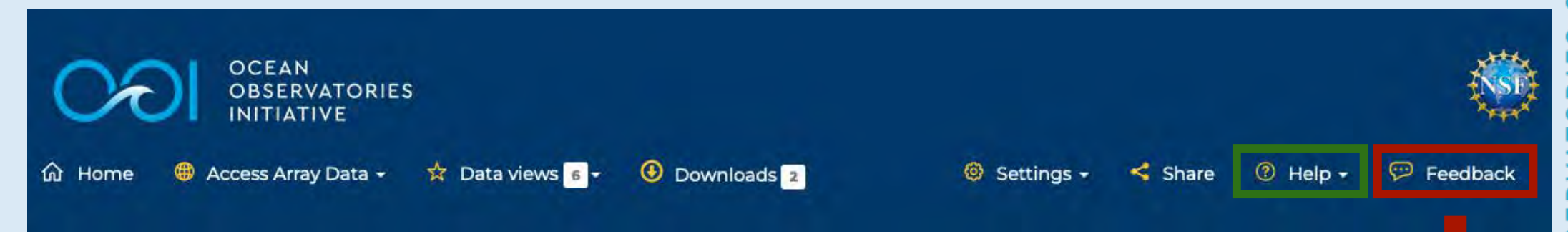
In [40]: def exists(path):
         r = requests.head(path)
         return r.status_code == requests.codes.ok

         def get_ooi_m2m(settings, url):
         try:
         r = requests.get(url, auth=HTTPBasicAuth(settings['M2M_USER'], settings['M2M_PASSWORD']))
         except Exception as ex:
         return ex
         json_returned = r.json()
         return json_returned
    
```


How to Get Help

- Data Explorer **Help** provides documentation and an interactive guide
- Data Explorer **Feedback** allows submission of text for feedback and/or issues
- Help and other topic are on the OOI Discourse site: <https://discourse.oceanobservatories.org/> This the recommended method to search, comment and post on topics for Help and others of interest such as Data Tools and Known Issues
- **OOINET** and **OOI** queries, comments or issues use helpdesk@oceanobservatories.org. Links can be found on OONET and the OOI main website

Data Explorer



Submit feedback

Comment or suggestion (required)

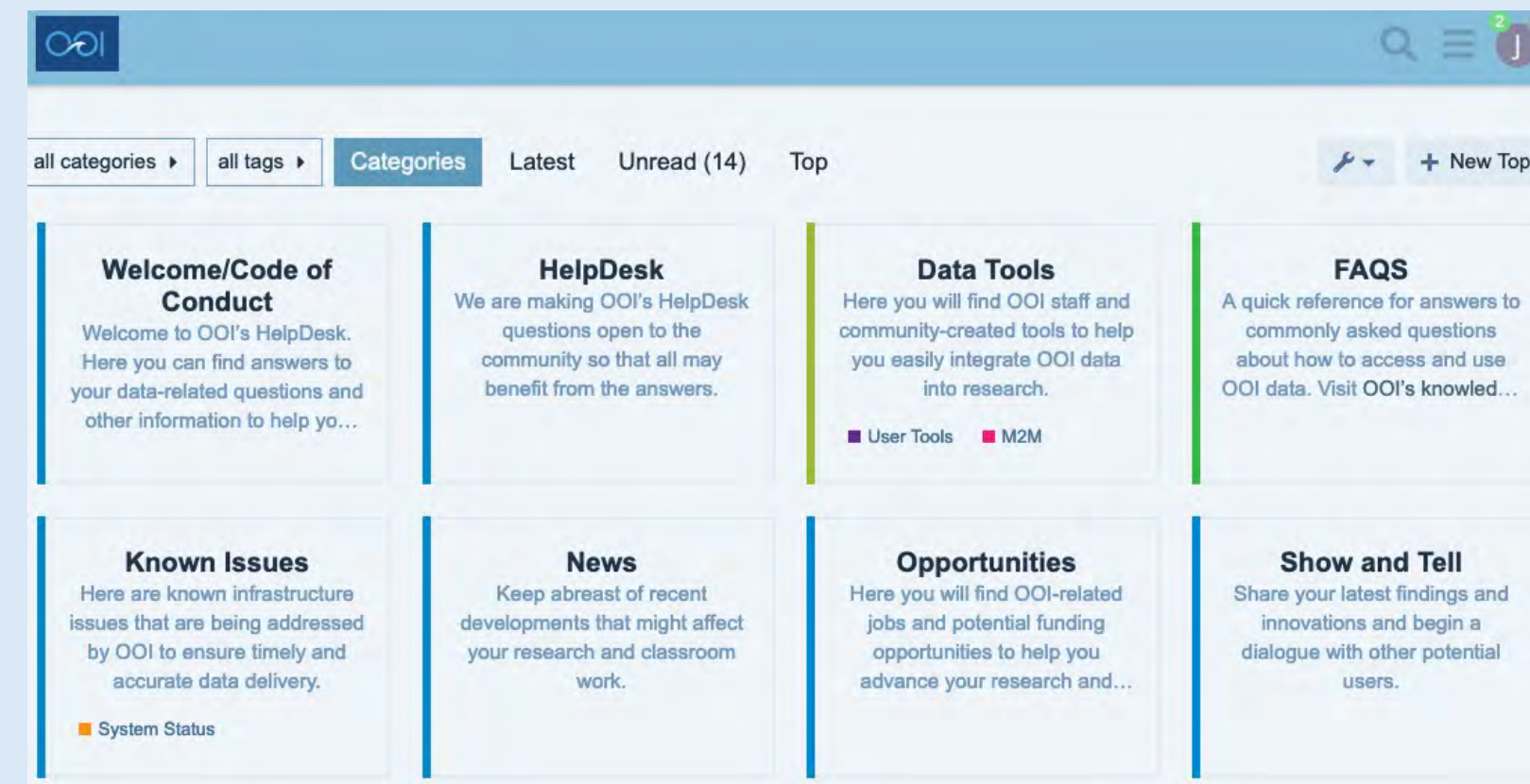
Your name

Your e-mail address

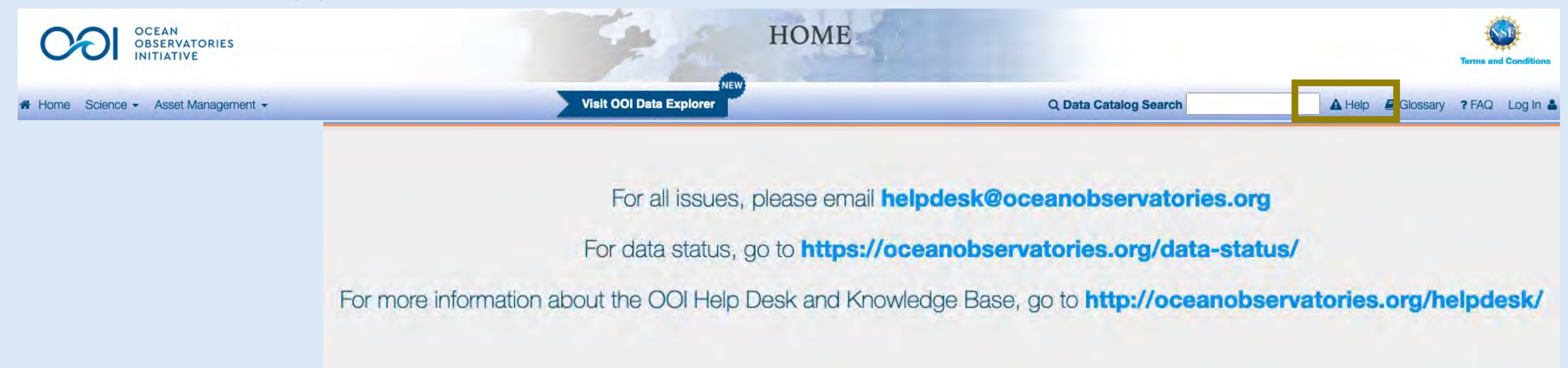
☐ I'm not a robot

Submit

Discourse site



OOINET



OOI Data Bus

More information and URLs can be found at: oceanobservatories.org/how-to-access-data/

Data Type / Source	Raw Data Archive	Thredds	Data Explorer	ERDDAP	M2M	OOI Website	Alfresco	Jupyter Hub	Data Explorer Media	OOINET
<div><div></div> Current best method</div> <div><div></div> Coming soon</div> <div><div></div> Phasing out</div>	Raw instrument and engineering datasets presented in an Apache file system structure for download.	Pre-computed scientific numerical data products with calibrations applied alongside engineering data. Full resolution datasets are accessible by deployment and stream.	Primary gateway to visualize and access OOI data. Search across data points, download full datasets using ERDDAP, compare datasets across regions and disciplines, and generate shareable custom data views.	Underlying data server for Data Explorer providing access to ~600 datasets organized by OOI arrays. Download datasets in common file formats and make graphs and maps.	Access to science and engineering data using both synchronous and asynchronous interfaces. NetCDF and JSON files are the standard outputs.	Oceanobservatories.org provides access to datasets compiled by Principal Investigators who have added instruments onto OOI arrays.	Document repository for instrument vendor information including calibrations	Hosted by OOI, this hub provides access to full resolution datasets and raw data server, allowing users to share notebooks and process data in a larger server environment.	Preview HD Photo, HD Video, visualized Hydrophone and ZPLS data along side science data in the Data Explorer tool	Legacy access to scientific and engineering data with the ability to search and plot data for review. Download requests are queued for system processing. User will be notified when download is ready. This interface will be slowly phased out.
Numerical Raw Data	<div></div>		<div></div>					<div></div>		
Processed Data Sets (NetCDF)		<div></div>	<div></div>	<div></div>	<div></div>			<div></div>		<div></div>
Provenance data (JSON)		<div></div>	<div></div>		<div></div>			<div></div>		<div></div>
Asset information						<div></div>	<div></div>			
Hydrophone	<div></div>							<div></div>	<div></div>	
Realtime data plots			<div></div>							<div></div>
ZPLS data	<div></div>							<div></div>	<div></div>	
PI Data						<div></div>				
HD Video	<div></div>							<div></div>	<div></div>	
HD Photo	<div></div>							<div></div>	<div></div>	

JupyterHub Demo

Presented by Will Koeppen



Notebooks authored by:

Axiom Data Science
Jim Case (Case Ocean Services)
Chris Wingard (OSU)

```
jupyter QA_Analysis Last Checkpoint: 11/24/2021 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)
+ -> Run C Code
# We have many files in the directory, but we want to subset them by time. In this example,
# we'll just subset by a start and end date.
start_day = datetime.datetime(2015,6,1,0,0)
end_day = datetime.datetime(2015,8,31,0,0)

# These lines build a datacube of time slices between our start and end dates,
# expanding the dimensions as it goes.

first = True

for i in range(nfiles):

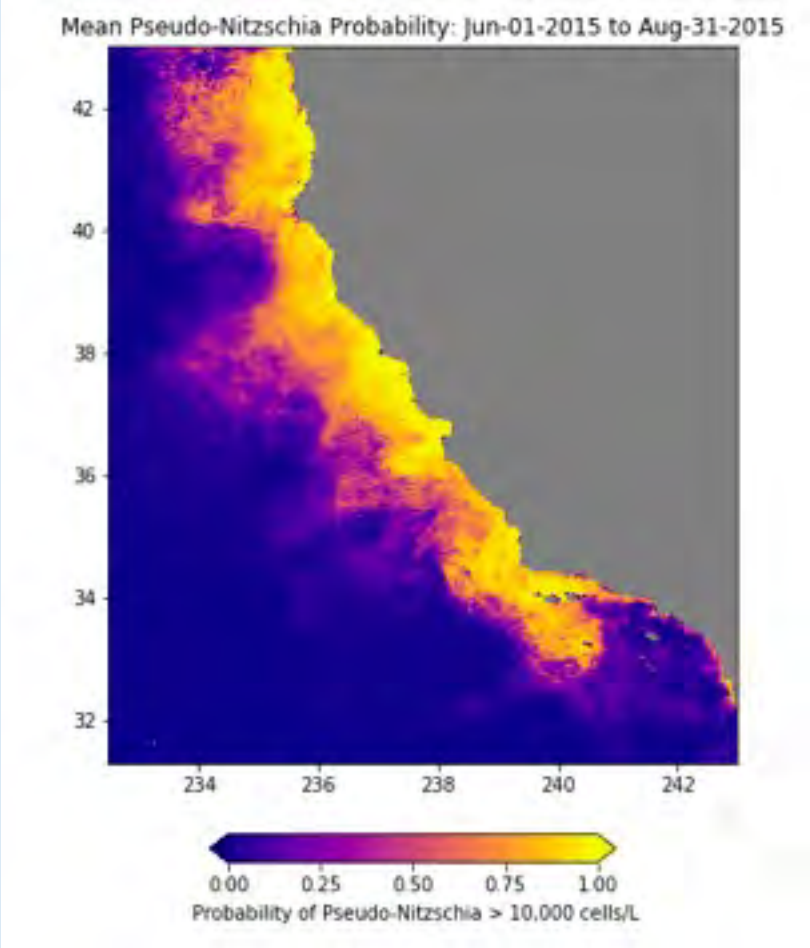
    netcdf = netCDF4.Dataset(filenames[i])
    # extract the time, turn it into a date
    t = np.array(netcdf.variables['time'])[0]
    t = netCDF4.num2date(t, time_var.units, time_var.calendar)

    # compare the date of the time slice to our set start and end dates
    if start_day <= t <= end_day:
        # get the data from the netcdf file, remove the first axis (time)
        thisdata = np.array(netcdf.variables[variable_name][0,t,:])

        if first:
            # If this is the first filename, create an array
            datacube = np.expand_dims(thisdata, axis=0)
            first = False
        else:
            thisdata=np.expand_dims(thisdata, axis=0)
            # If this is not the first filename, add to the existing array
            datacube= np.append(datacube, thisdata, axis=0)
    # close each file that we open
    netcdf.close()

100% [ ] 1222/1222 [00:24<00:00, 50.09it/s]

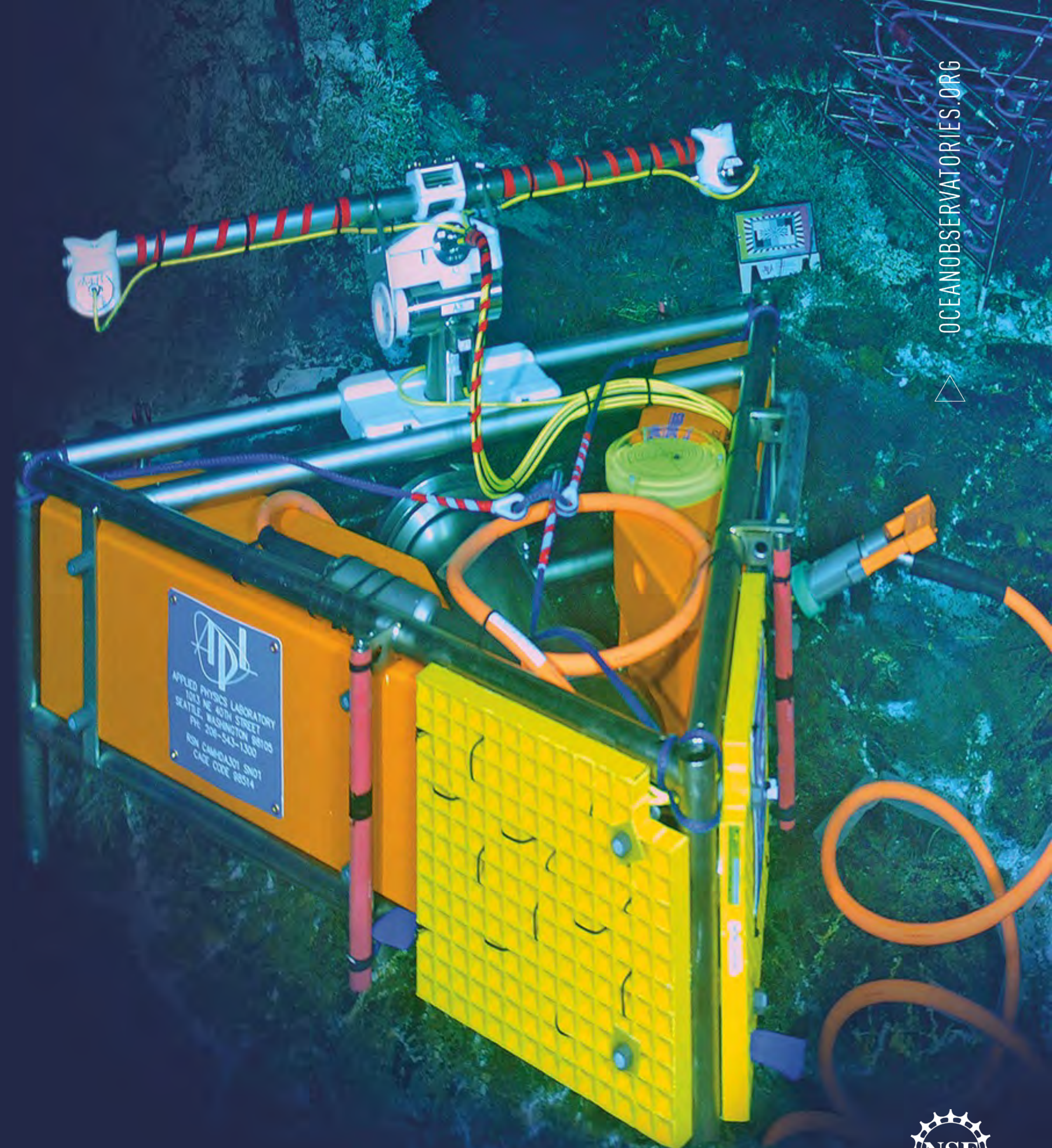
# Apply the mask to the datacube
datacube = na.masked_values(datacube, -999.)
```





OCEAN
OBSERVATORIES
INITIATIVE

Questions?



OCEANOBSERVATORIES.ORG

