

Using OOI Data Programmatically with Pangeo

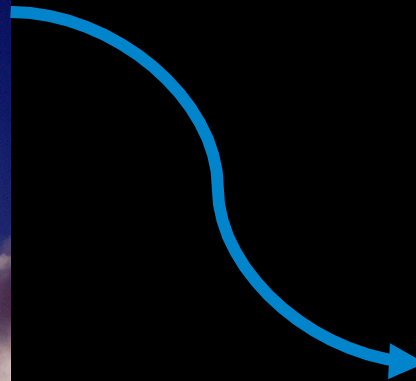
Rich Signell
USGS, Woods Hole, MA

NSF DDCI
2020-10-27

Traditional Model Data Analysis



Model Data Analysis of the Future (available now!)



Pangeo is a Community



PANGEO

A community platform for Big Data geoscience

Pangeo is an open group. Anyone who agrees with our [mission and vision](#) is welcome to join.

To add your name to the list, fork the [source for this site](#) on GitHub, add your details to the file `_data/people.yml`, and submit a pull request. The easiest way to do this is to directly [edit the file](#) on GitHub.

Ryan Abernathey



Lamont Doherty Earth Observatory
physical oceanography climate

How fast can the Met Office's solution pull data from S3?
#198 opened 17 days ago by mrocklin

16

Pangeo use case: Advanced regridding using ESMF/ESMPy/OGGIS/xESMF/Xarray/Dask
#197 opened 17 days ago by jhamman

18



Pangeo is a Flexible Open-Source Framework



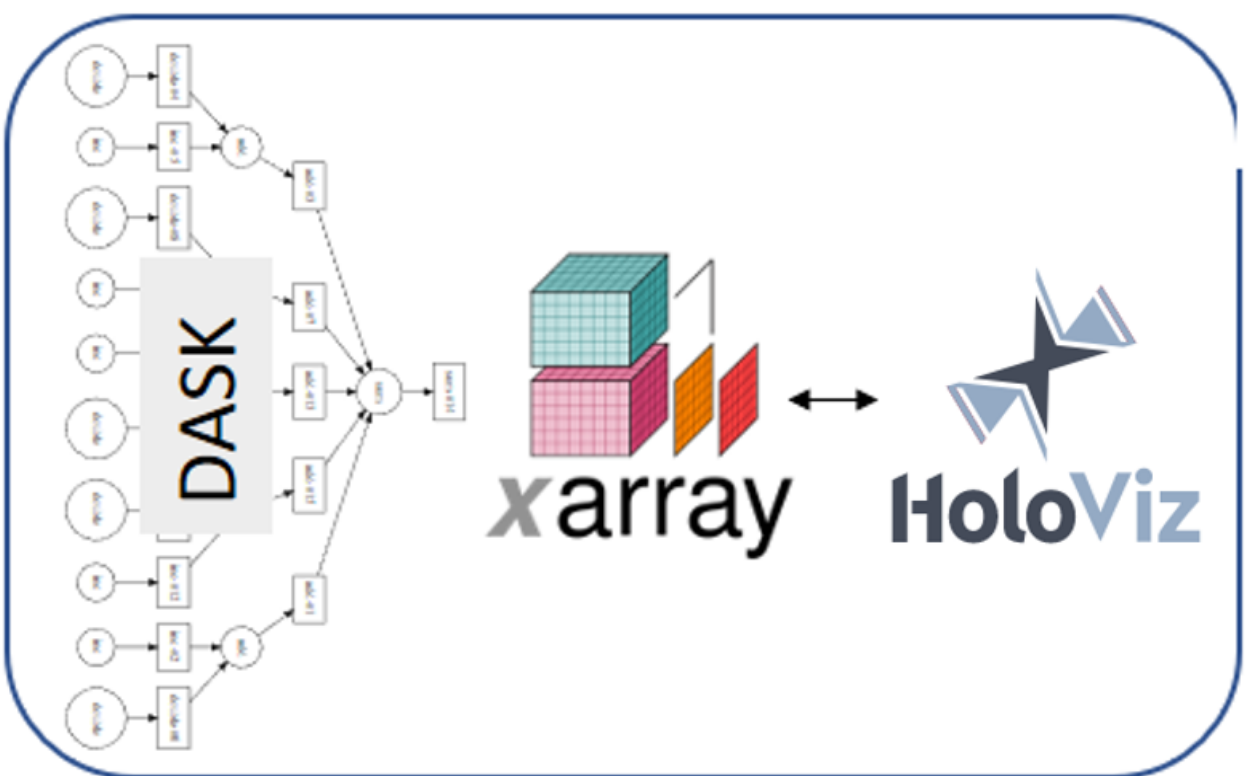
Credit: Stephan Hoyer, Jake Vanderplas (SciPy 2015)

Pangeo Cloud Architecture

Cloud-friendly ndarray data



Cloud Analysis and Visualization Environment



User with Browser



Cluster = KubeCluster() or GatewayCluster()

SLURMCluster

Workers 20

► Manual Scaling

Cores 20

► Adaptive Scaling

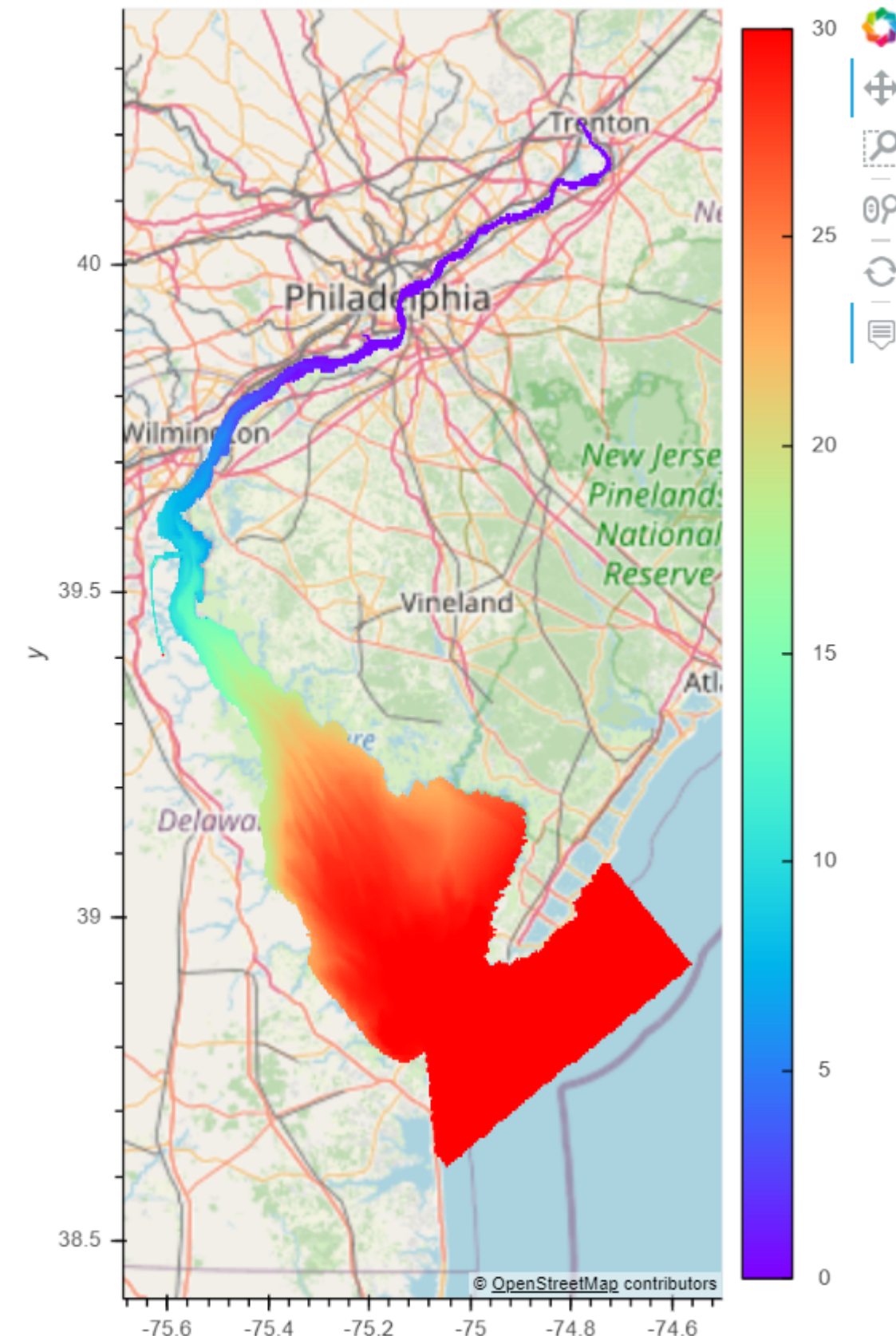
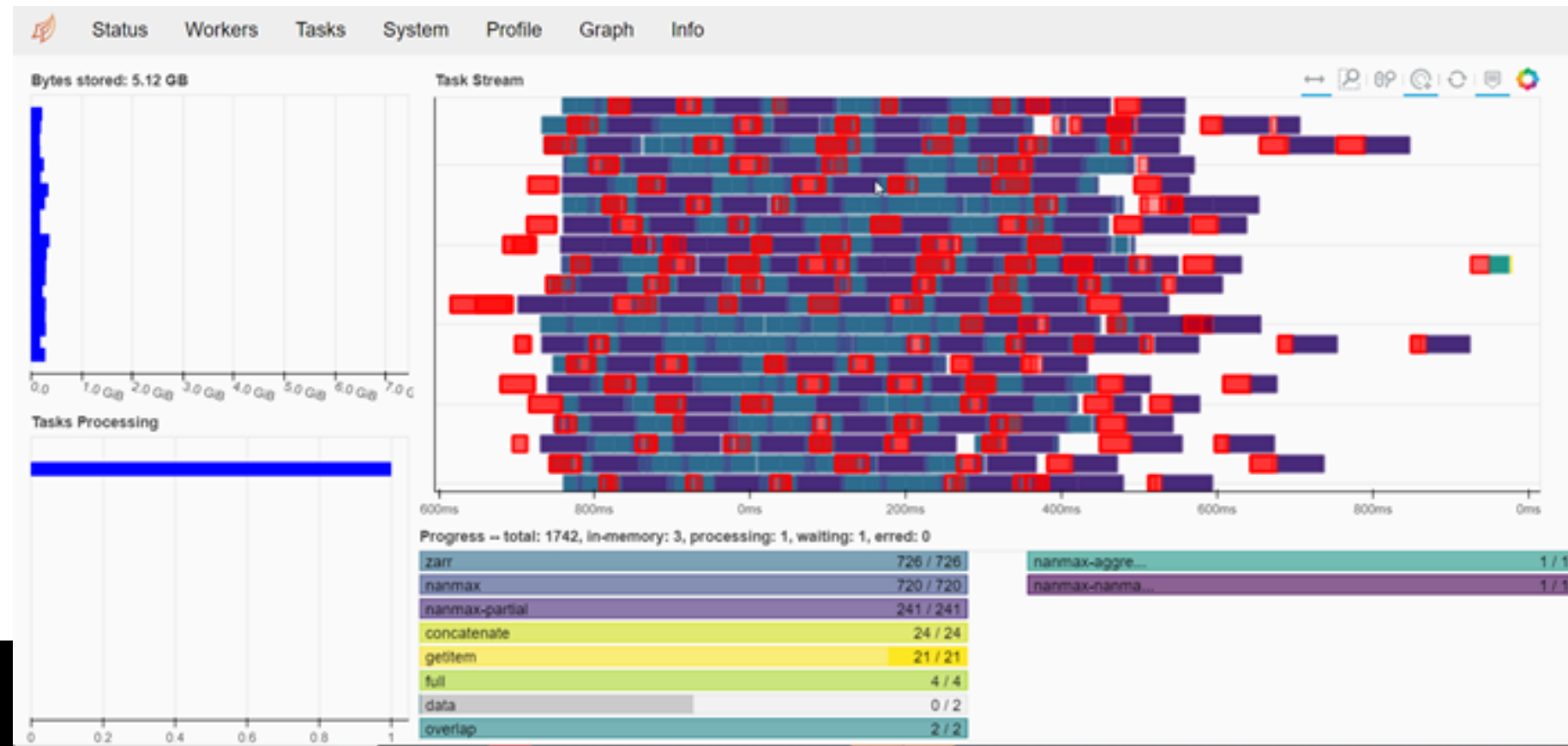
Memory 160.00 GB

Dashboard: <http://10.128.0.39:8787/status>


```
[35]: %%time
ds.salt.max(dim='ocean_time')[-1,:,:].hvplot.quadmesh(x='lon_rho',y='lat_rho',
    geo=True, tiles='OSM', rasterize=True, cmap='rainbow', frame_width=300)
```

CPU times: user 38.7 ms, sys: 0 ns, total: 38.7 ms

Wall time: 36.4 ms




Pangeo is not just for model data...



ABOUT PANGEO | WEBSITE

Cloud Native Geoprocessing of Earth Observation Satellite Data with Pangeo



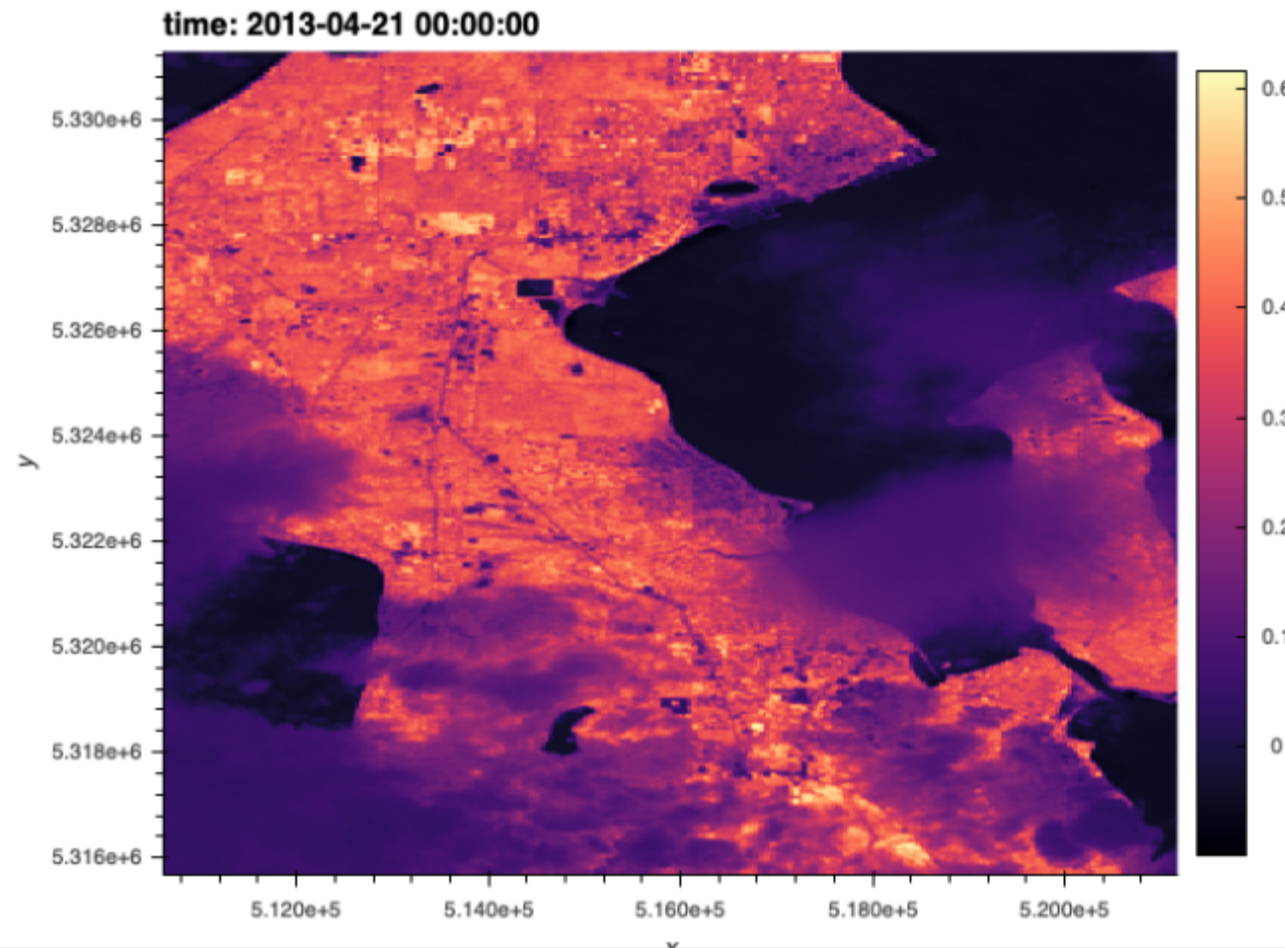
Scott Henderson

Oct 1, 2018 · 7 min read

If you are familiar with satellite imagery you've likely heard that we are entering a "golden era" of Earth Observation. It's true! New satellites are generating Petabyte-scale publicly available archives of imagery at unprecedented rates, enabling new insights and fast global impacts.

Add a scott s
perhaps from

Cloud Native Landsat Analysis with Pangeo



We've developed an example Cloud Native quantitative analysis of Landsat 8 satellite imagery. What is special about this example is that the analysis is easily reproduced, scalable, and interactive: 100 Gigabytes of Landsat 8 images covering Washington State (representing the entire archive back to 2013-03-21) are found using NASA's Common Metadata Repository (CMR). Then, using URLs instead of local file paths, the Normalized Difference Vegetation Index (NDVI), a simple landcover

Pangeo is not just for geoscience data

PanNeuro: leveraging a community-based approach for big data neuroscience

BRAIN Initiative PI meeting, April, 2019

Ariel Rokem

The University of Washington eScience Institute

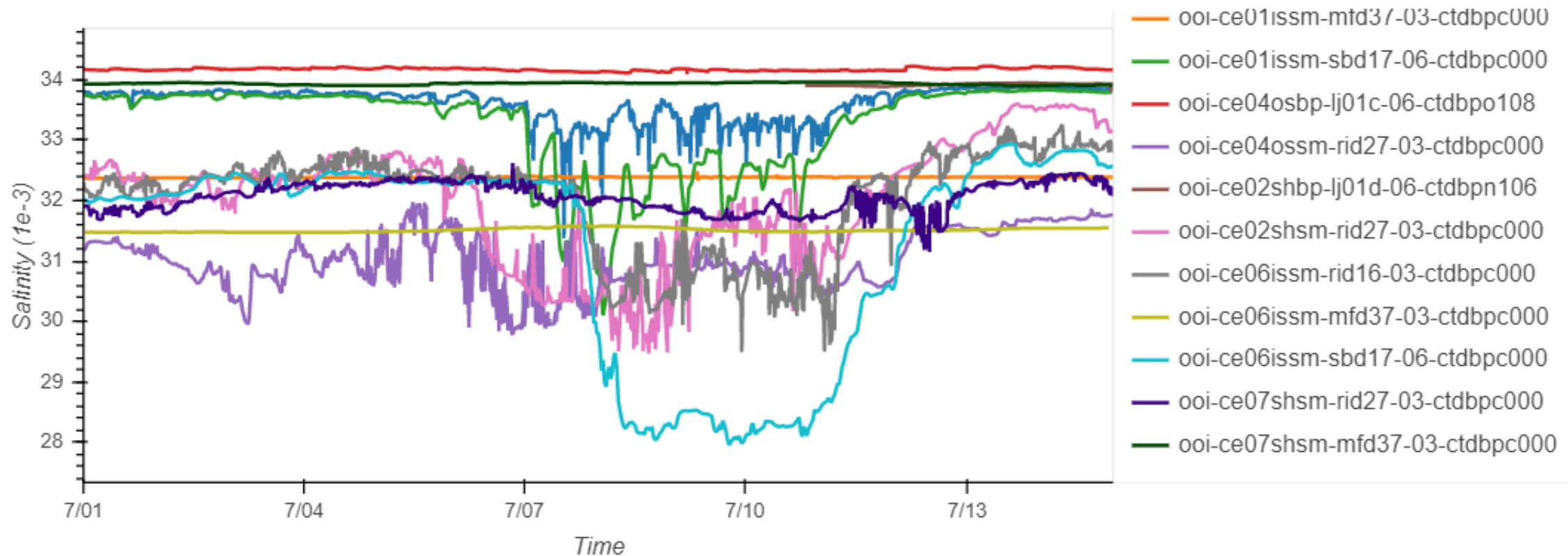
Follow along at: <https://arokem.github.io/2019-BRAINI-PanNeuro-slides/>



Pangeo is also useful for exploring OOI data

```
In [17]: hv.Overlay(hv_list).opts(legend_position='right', width=900, legend_offset=(0,0))
```

Out[17]:



See and run the [ERDDAP Access Demo](#) from the Pangeo Gallery!

Summary

- Pangeo is a flexible open-source framework for scalable, data-proximate data access
- The new OOI ERDDAP provides a powerful way to programmatically search and access OOI data
- Future Wish: a standard access layer: “gold standard” NetCDF files on Cloud storage, to access OOI data directly without services