PRESENTER: **Roland Schweitzer**

INTRO

- There are hundreds of millions of real-time ocean observations representing years of data collection locked up in a proprietary database. While these observations are available via ERDDAP backed by a special database view, access is slow and limited. This experiment tests the viability of using Google Big Query and other cloud databases to store the data and accessing it directly with it via the API or through an ERDDAP via JDBC.

METHODS

1. To keep from overwhelming the current service we transferred data one day at a time (about a million records) with a pause between each transfer.
2. Once that data were loaded, we configured an ERDDAP to connect to the service via JDBC.
3. We built a dashboard with Plotly Dash which can be used to explore the data.
4. We hope to repeat this with other cloud platforms.
5. Once the technology is proven we will consider submitting the data for inclusion in the NOAA Open Data Dissemination (NODD)

RESULTS

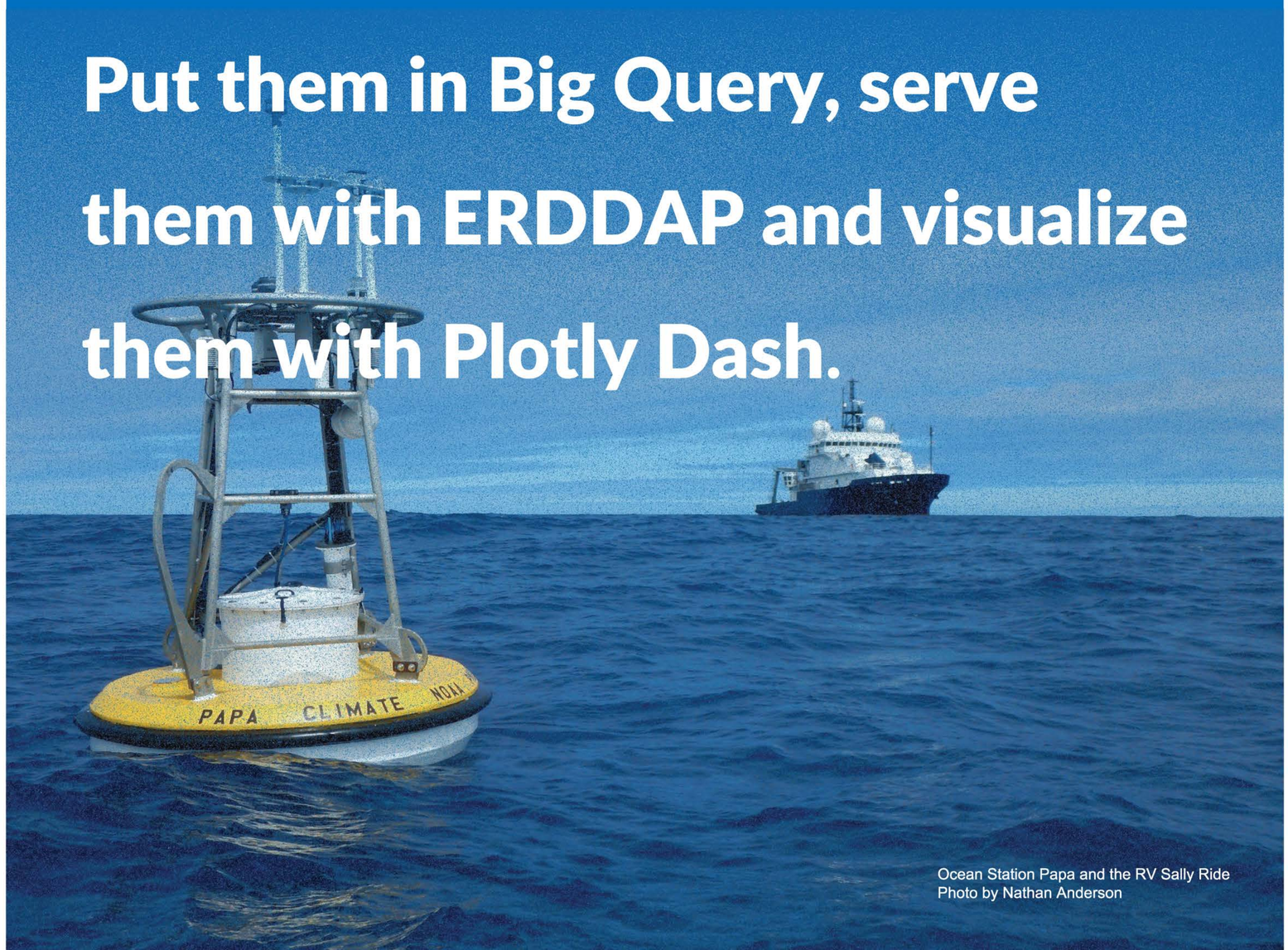
- Query time via the Big Query API is good. The data transfer time scales with the size, but visualizing the data in 3-month chunks works well.
- Access via ERDDAP is still pretty slow, but can take advantage of ERDDAP decimation techniques.

DISCUSSION

- This technique combined with NOAA NODD might open up new avenues for sharing ocean observations.

What do you do with hundreds of millions of ocean observations?

Put them in Big Query, serve them with ERDDAP and visualize them with Plotly Dash.

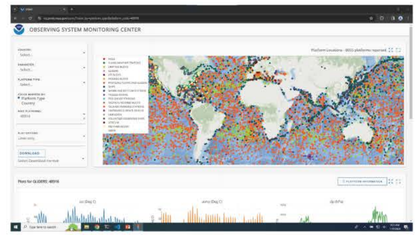


Ocean Station Papa and the RV Sally Ride
Photo by Nathan Anderson

Pacific Marine Environmental Laboratory

50 years of Science in Service to Society

Roland Schweitzer
Kevin O'Brien
Eugene Burger



We will release a data exploration dashboard for these data soon.

We have also compiled a Big Query table which contains the known metadata for the observing platforms which is linked to the data via the platform ID.

There are 35 different columns of data though most platforms are reporting on a limited number of parameters.

At the time of this presentation the Big Query table contained 151 million rows representing data from 1 Jan 2020 to 29 August 2020.

This work is sponsored by NOAA's Global Ocean Monitoring and Observing Program.

Pacific Marine Environmental Laboratory