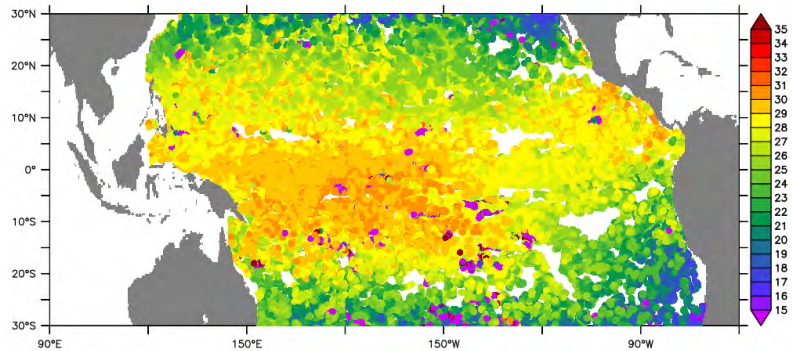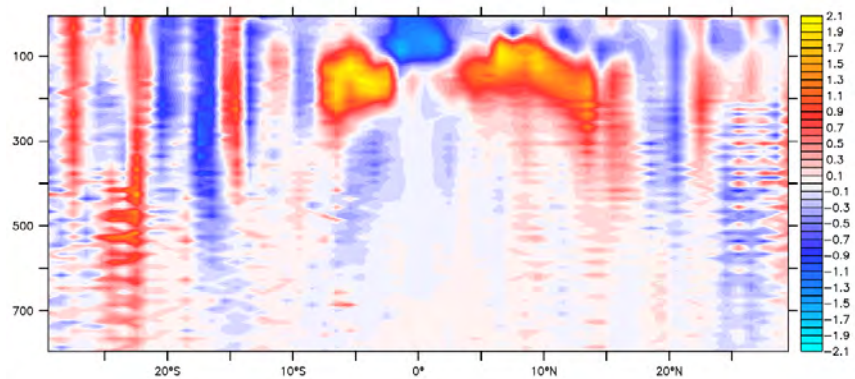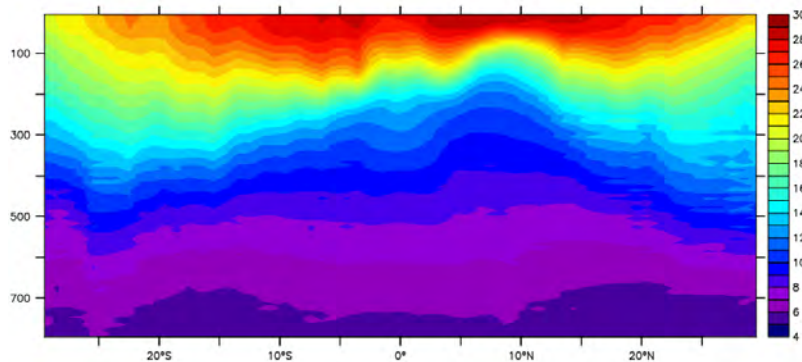# Integrating Data and Information Across Observing Systems

Kevin O'Brien: University of Washington, WA, USA
Steven Worley: UCAR, CO, USA
Bob Simons: NOAA Southwest Fisheries, CA, USA
Eugene Burger: NOAA PMEL, WA, USA

Benjamin Pfeil
Bjerknes Climate Data Centre
University of Bergen, Norway

Integration and Interoperability, what is it?

Three broad categories of integration by example:

1. Collecting single observing system observations
   - All data from the Surface Ocean CO_2 Atlas (SOCAT) project in a common data and metadata format
2. Multiple observing systems, across an EOV/ECV (Temperature)
   - Surface Temp from moored and drifting buoys, ARGO, and hydrographic stations
3. Cross-system, multi-variable reference collections
   - World Ocean Database (WOD), International Comprehensive Ocean-Atmosphere Data Set (ICOADS), Surface Ocean $C0_2$ Atlas (SOCAT)

Interoperability – definition[1]

The ability of data or tools from <u>non-cooperating resources to integrate</u> or work together with minimal effort.

[1] Wilkinson, M. D. et al. *The FAIR Guiding Principles for scientific data management and stewardship.* Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

FAIR = Findable, Accessible, Interoperable, Reusable

- Example – Expedia.com: "Find me a cheap flight"
  - Poll many systems in real time
  - Using known protocols, i.e. access mechanism

Surface Ocean CO2 Atlas (SOCAT) v3

traditi...

Index of /ftp/oceans/SOCATv3

cdiac.ornl.gov/ftp/oceans/SOCATv3/

Name	Last modified	Size	Description
Parent Directory			
Fair_Data_Use_Statement_for_SOCAT.pdf	29-Jan-2016 15:35	85K	
Read_SOCATv3.m	15-Sep-2015 08:21	18K	
SOCAT_cat_logo.jpg	25-Aug-2015 15:10	835K	
SOCATv3.zip	29-Oct-2015 11:27	330M	
SOCATv3_Arctic.zip	29-Oct-2015 11:20	14M	
SOCATv3_Coastal.zip	29-Oct-2015 11:23	127M	
SOCATv3_FlagE.zip	29-Oct-2015 11:29	5.3M	
SOCATv3_Gridded_Dat/	02-Mar-2016 09:22	-	
SOCATv3_Indian.zip	29-Oct-2015 11:20	4.2M	
SOCATv3_NorthAtlantic.zip	29-Oct-2015 11:21	35M	
SOCATv3_NorthPacific.zip	29-Oct-2015 11:22	13M	

Or available via FTP grouped by ocean basin

Though downloadable, the user will have to:
- Download all 4K files
- Somehow knit them together in order to look at the entire "collection"
- Format the data to work in application of choice

Specifically, in regional studies invite large data providers, who frequently possess valuable expert knowledge on data and region, to collaborate at an e which may lead to an invitation of co-authorship.

We recognize that co-authorship is only justified in case of a significant scientific contribution to a publication and that provision of data on its own warrant co-authorship.

2) Cite SOCAT and its data products as:
Version 3: Bakker et al. (2016, in prep.) and Bakker et al. (2014) (cite both papers until the version 3 publication has been published in ESSD);

Surface Ocean CO2 Atlas (SOCAT) v3 *interoperable* data access

Because data are available through ERDDAP framework
- No Download necessary (DAP access avaialable)
- User can look at whole "collection" of data as a single dataset
- Many applications can directly access and use data from ERDDAP
- Easy to leverage existing tools to visualize and analyze data

Surface Ocean CO2 Atlas (SOCAT) v3 interoperable data access
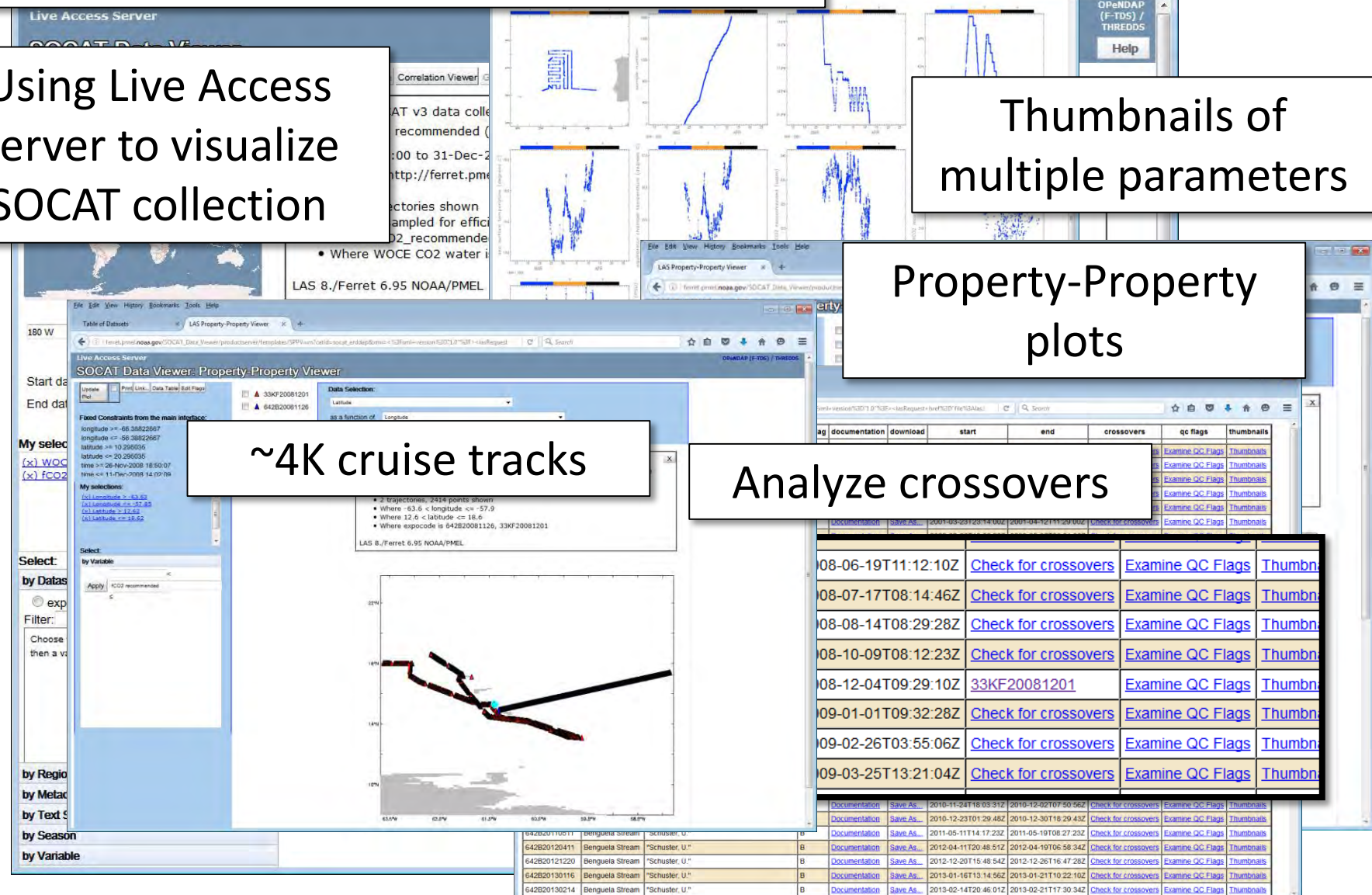
Using Live Access Server to visualize SOCAT collection

Thumbnails of multiple parameters

Property-Property plots

~4K cruise tracks

Analyze crossovers

Other clients (Excel, Google Earth, Matlab®, Jupyter, Notebooks, R, etc) can also directly access data from ERDDAP

Digression: DAP in the *modeling* community

Data Access Protocol (DAP) (originally Distributed Oceanographic Data System (DODS)), revolutionized how data was accessed
- User could use familiar applications and not worry about data formats
- Users could access complete model runs with a single command (time series analysis, etc)
- Allows for access of heterogeneous datasets in a highly distributed environment

OPeNDAP servers (Hyrax, THREDDS data server) are now the de facto way to serve such data.

Many barriers to interoperable access of gridded/model data have been removed thanks to DAP
- See Unified Access Framework project which offers access to 6K+ datasets in a uniform way

Digression: DAP for the ocean observation community

ERDDAP is DAP, but for in situ data!

- ERDDAP can also revolutionize how people serve, access and use observational data
- ERDDAP is NOT a portal - it's a framework upon which portals can easily be built
- From Provider standpoint:
  - Can ingest many existing formats and make available through interoperable web services – flexible!
- From user standpoint
  - Can access and use observations and "collections" of observations with familiar clients without worrying about data format
  - Provides machine to machine access with RESTful web services

ERDDAP is DAP, but for in situ data!

- Reducing the time required to deal with data management allows more time for scientific productivity

- Does it solve all the problems?
    - No. But can significantly improve integration of many types of data already.
    - Leveraging existing standards, conventions and frameworks, such as ERDDAP, allows us to solve many of these issues at significantly reduced costs.

JCOMM Observations Coordination Group (OCG) ACTION: to develop vision for integration of data across observing systems.

A Piloting (prototyping) Activity to Address Key Challenges of the JCOMM OCG Task Team on Integrated Data Access

- Chose to focus on Tropical Pacific in support of TPOS 2020
- Build an integrated framework of ocean observation data from multiple platform networks or data sources:
    - Drifting Buoys
    - Argo Floats
    - SOT/VOS data (CCHDO)
    - SOCAT underway data
    - IOOS Glider data
    - OceanSITES stations
        - Tropical Moorings, Stratus, WHOTS, CCE1,CCE2
    - Real Time data from OSMC GTS
- Selected ERDDAP as data framework for integrating the observations

JCOMM Observations Coordination Group (OCG) ACTION: to develop vision for integration of data across observing systems.

Further possibilities

- Integrating across EOV – Temperature example

JCOMM OCG ACTION: to develop vision for integration of data across observing systems.

All Data collected and available through a single ERDDAP server

ERDDAP, we can create an "integrated" view of upper ocean temperature

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	WMS	Source Data Files	Title									
	set	data	graph			* The List of All Active Datasets in this ERDDAP *								al
	set	data	graph			AOML Interpolated Drifter Observations								dr
	set	data	graph			Argo float vertical profile (ARGO delayed mode data)								pa
	set	data	graph			CCHDO CTD FILES								ct
	set	data	graph			CCHDO Hydrographic data								hy
data			graph	M		Gridded upper ocean temperature from delayed mode sources								gr
	set	data	graph			IOOS Pacific Glider data								pa
data			graph	M		March Real Time Gridded Temperature data	❓	F I M	background ☑	RSS	✉	OSMC		gr
	set	data	graph			OceanSITES CCE1 Mooring data	❓	F I M	background ☑	RSS	✉	OceanSITES/UCSD/SIO	C	
	set	data	graph			OceanSITES CCE2 Mooring data from the California Current	❓	F I M	background ☑	RSS	✉	USCD		m
	set	data	graph			OceanSITES Mooring Temperature (TAO Delayed/Mixed Mode 10 min Data)	❓	F I M	background ☑	RSS	✉	oceansites		m
set	data	graph				OSMC Real Time GTS Ocean observations (OSMC flattened observations from GTS)	❓	F I M						st
set	data	graph				OSMC Upper Ocean Temperature from the Pacific	❓	F I M						in
set	data	graph				SOCAT v3 data collection	❓	F I M						sc
	set	data	graph			SOCAT v3 data collection	❓	F I M	background ☑	RSS	✉	PMEL		sc
	set	data	graph			TAO Delayed/Mixed Mode 10 minute Data (OceanSITES Mooring MET data)	❓	F I M	background ☑	RSS	✉	OceanSITES		m
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, ADCP	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, Air Temperature	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, Precipitation	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, Salinity	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, Sea Surface Salinity	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, Sea Surface Temperature	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	
	set	data	graph			TAO/TRITON, RAMA, and PIRATA Buoys, Daily, Temperature	❓	F I M	background ☑	RSS	✉	NOAA PMEL, TAO/ ... ❓	pr	

use http://ferret.pmel.noaa.gov/generic/erddap/tabledap/integrated_temp.nc
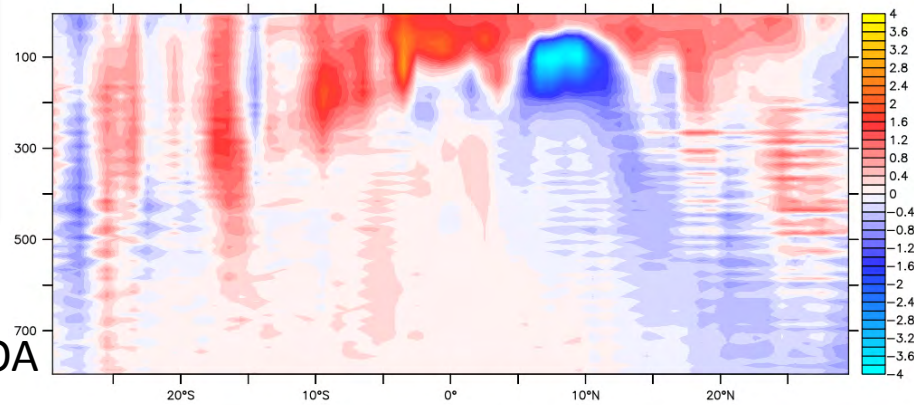
shade/z=0 temp

As a result, rather than a data management project, users can easily access an integrated view of upper ocean temperature with one command
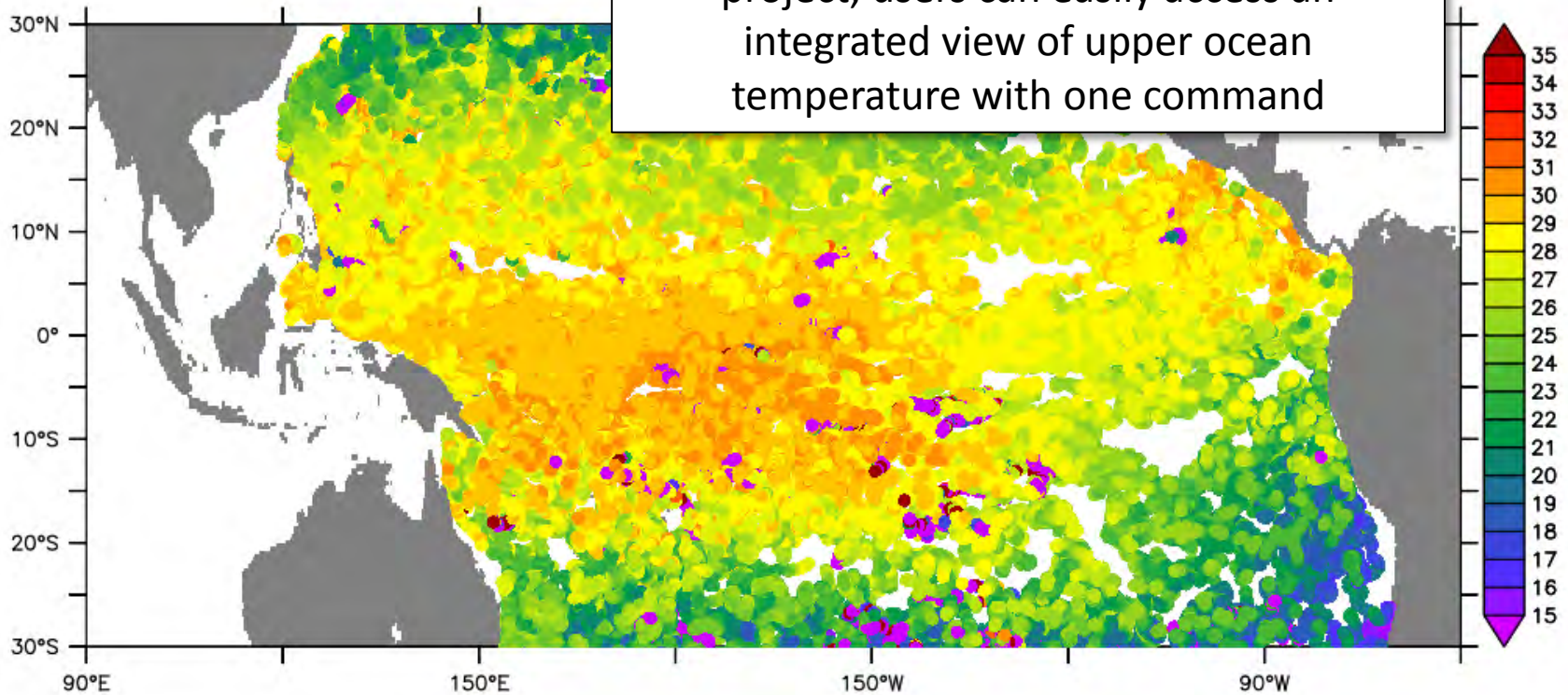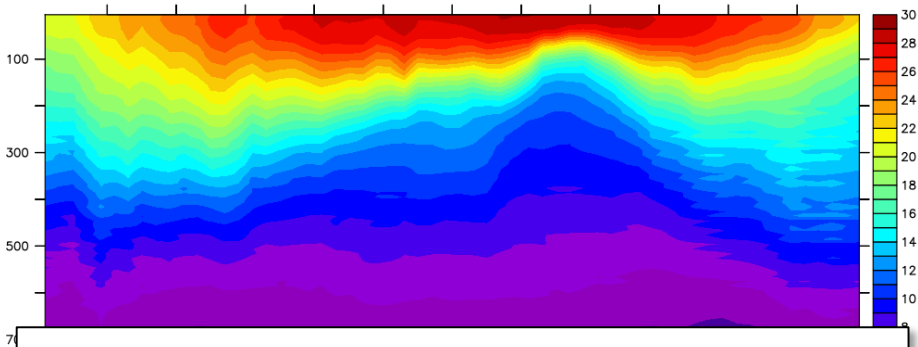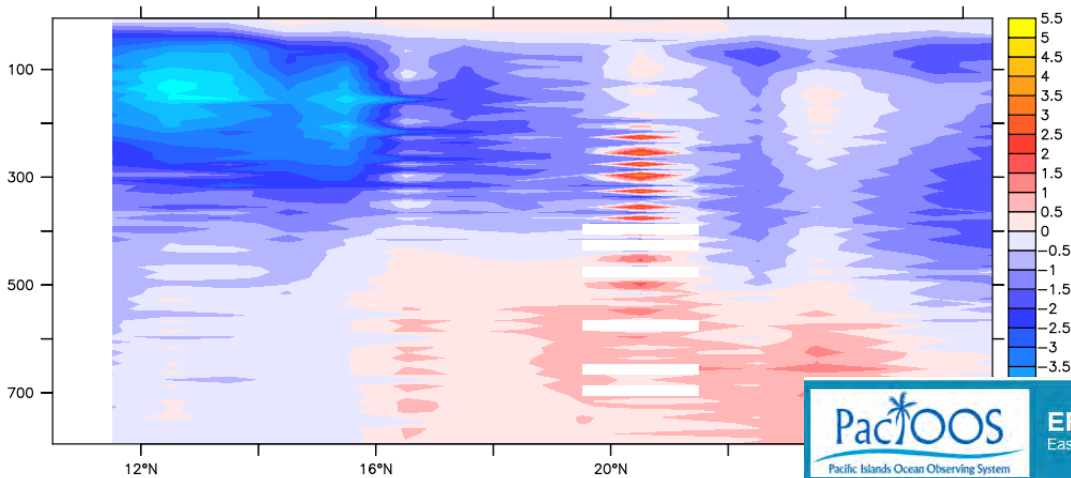
2015 Tropical Pacific temperature, averaged over time and longitude

And start using the data for analysis in their favorite application

2015 TEMP Anomalies from WOA

Oct 2015 TEMP Anomalies from PacIOOS ROMS CNMI

PacIOOS
Pacific Islands Ocean Observing System

ERDDAP
Easier access to scientific data

Moving Forward…..

JCOMM Data Management Coordination Group (DMCG) recommendations:

- Observation Coordination Group continue working with global observing networks to implement data interoperabilty framework in close consultation with DMCG

- Welcomed proposed OCG pilot project to demonstrate integration by EOV

- Requested OCG refine the vision of EOV pilot project and consult with observing networks to determine interest

- OCG to engage with GOOS Regional Alliances (GRA) to provide data framework in support of GRA data portals and ODP

Thank you!

Kevin O'Brien: Kevin.M.O'Brien@noaa.gov

Benjamin Pfeil: benjamin.pfeil@uib.no

- ERDDAP: https://coastwatch.pfeg.noaa.gov/erddap/index.html
- OSMC: www.osmc.noaa.gov

Poster 102: Global Data Assembly Centre for Marine Biogeochemistry will be build services on ERDDAP!