Moving Towards FAIR Data Principles with ERDDAP

KEVIN O'BRIEN

UNIVESITY OF WASHINGTON/JISAO

NOAA/PMEL

OBSERVATION COORDINATION GROUP VICE-CHAIR FOR DATA AND INFORMATION

NSF EarthCube Workshop for Ocean Time Series Data, September 13-15, 2019

GOOS Observation Coordination Group (OCG)

- OCG coordinates activities of the global ocean observing networks
- OCG is involved with near-time delivery of data as well as delayed mode data access
- OCG is working to improve data interoperability between and within the various observing networks.
- OCG engages at global, regional and local levels to provide a common set of data services.

All data contains temp, but....

Drifting Buoys — ASCII files

Argo Floats — NetCDF

OSMC GTS — Oracle database

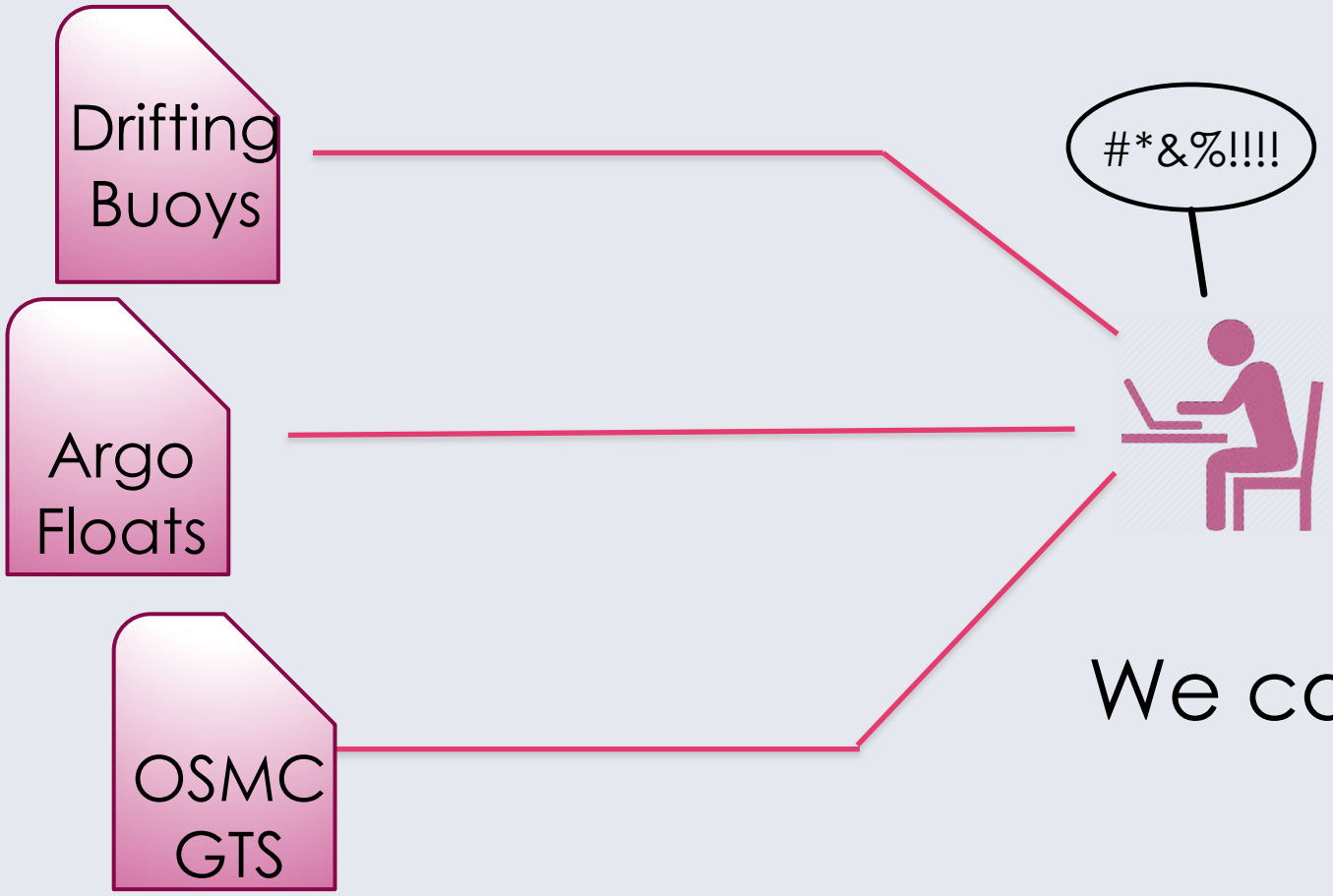
All data contains temp, but....

Drifting Buoys — ASCII files

Argo Floats — NetCDF

OSMC GTS — Oracle database

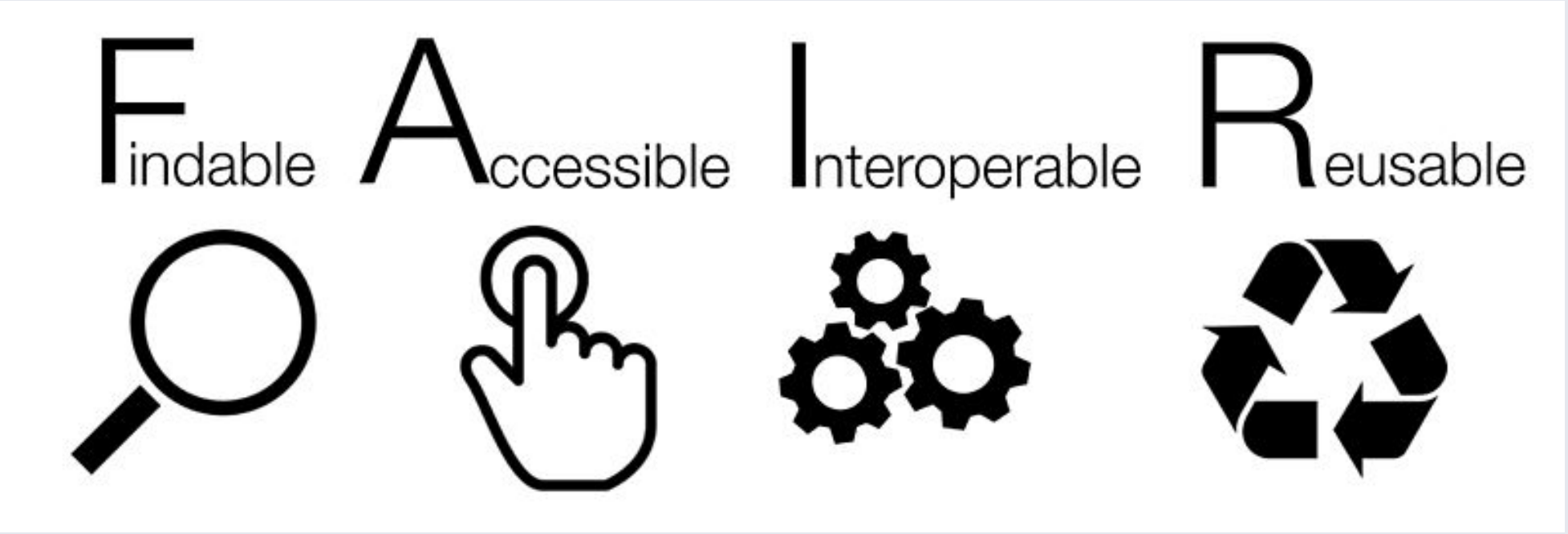
All data contains temp, but....



Three different formats means:

- Three different download methods
- Have to convert two or three of the data streams
- Suddenly, this is a data project!!

We can do better than this!



FAIR data principles will help frame the forthcoming OCG data strategy

- FAIR data principles provide an excellent rallying point
- Focus is on all digital – not specifically Ocean data
- Abstract concepts can be hard to make actionable
- Need to specify pragmatic ways to move towards FAIR compliance for OUR communities
- Data Stewardship needs to go beyond FAIR as well

ERDDAP

Why ERDDAP

- A data brokerage service, reading from many different types of files, databases and services, and providing access via a single standardized interface (interoperability layer)
- RESTful API for access in scientific analysis packages (Matlab, Python, R), web application developers (JavaScript), and by numerical modelers (Fortran, Bash)
- Advanced search built-in, and also generates ISO and json-ld metadata records to allow search via sites like data.gov, and Google Dataset Search.
- Widely used for delivery of "FAIR" data in the geoscience community (many server deployments worldwide)

Why ERDDAP

BENEFITS FOR PROVIDERS

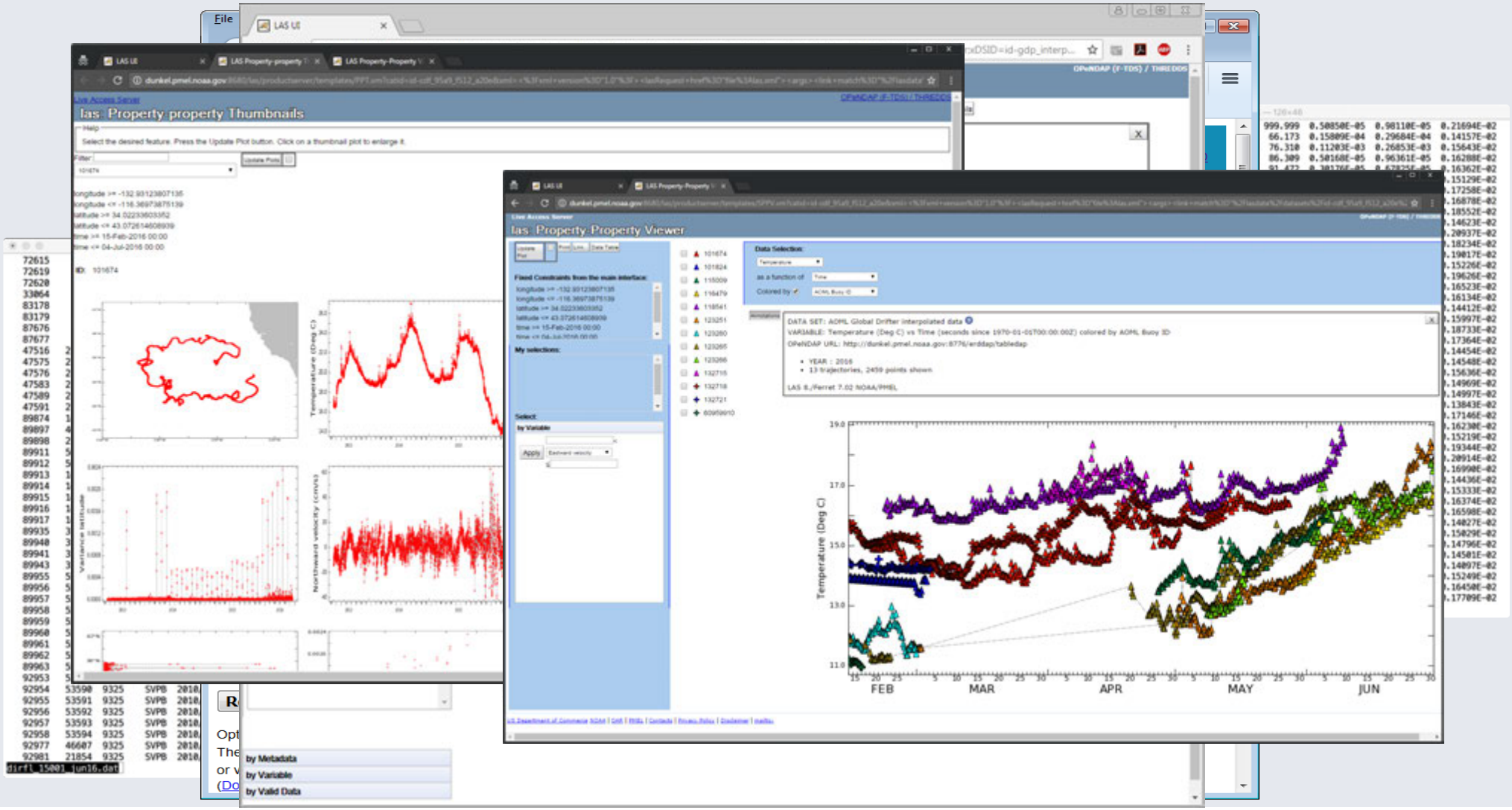
- Provides data to users in **many different formats** and m2m services
- Can augment metadata without rewriting files
- Can automatically create BagIt documentation files for archival purposes

BENEFITS FOR USERS

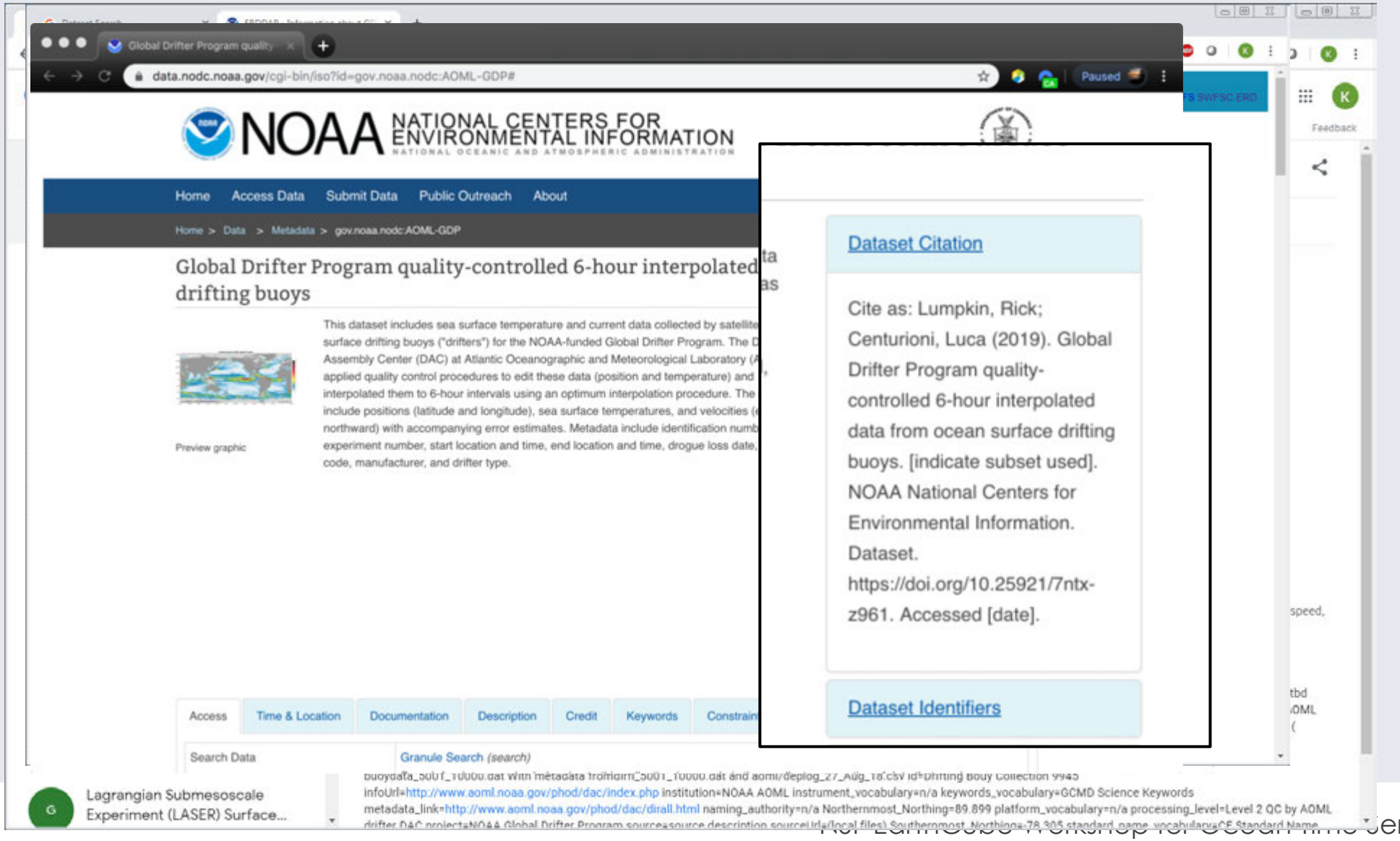
- Users can access data with clients of their choice without reformatting data
- Supports m2m capabilities for building community specific access services
- Can access "collections" of aggregations

AOML Global Drifter Program

JISAO PMEL Pacific Marine Environmental Laboratory



AOML Global Drifter Program



ERDDAP (cont.)

Where it's used

- Surface Ocean CO_2 Atlas (SOCAT) project
- OceanSITES – serving long timeseries data
- Argo, Global Drifters, Ocean Glider DACs, Sea Level (in progress), Animal Sensors (in progress)

V2 features

- ERDDAP data ingest capability (insert data via a URL)
  - Useful for automatically loading data into ERDDAP (sensor data)
- Enhanced server-side filtering operations for constraining data requests
- Web-page ready data formats (for google chart, etc.)

ERDDAP (cont.)

Where it's used

- Surface Ocean CO_2 Atlas (SOCAT) project
- OceanSITES – serving long timeseries data
- Argo, Global Drifters, Ocean Glider DACs, Sea Level (in progress), Animal Sensors (in progress)

V2 features

- ERDDAP data ingest capability (insert data via a URL)
  - Useful for automatically loading data into ERDDAP (sensor data)
- **Enhanced server-side filtering operations for constraining data requests**
- **Web-page ready data formats (for google chart, etc.)**

ERDDAP services and FAIR principles

F

- ERDDAP can easily augment metadata with missing elements for compliance
- ERDDAP can generate ISO metadata documents on demand (harvestable)
- ERDDAP supports schema.org metadata

A

- ERDDAP provides access to data in multiple formats
- ERDDAP supports RESTful services for machine to machine access (harvestable)
- ERDDAP provides a platform upon which data-specific applications can be built

I

- All of these capabilities lead to improved interoperability

R

- ERDDAP supports creation of BagIt packages for easy archival (DOI)
- ERDDAP supports Climate and Forecast metadata, json-ld, schema.org

Thank you!

ERDDAP:
https://coastwatch.pfeg.noaa.gov/erddap/

Awesome ERDDAP (list of ERDDAPs and ERDDAP applications):
https://github.com/IrishMarineInstitute/awesome-erddap

Kevin.M.O'Brien@noaa.gov, kob@uw.edu
Bob.Simons@noaa.gov

NSF EarthCube Workshop
for Ocean Time Series Data,
September 13-15, 2019