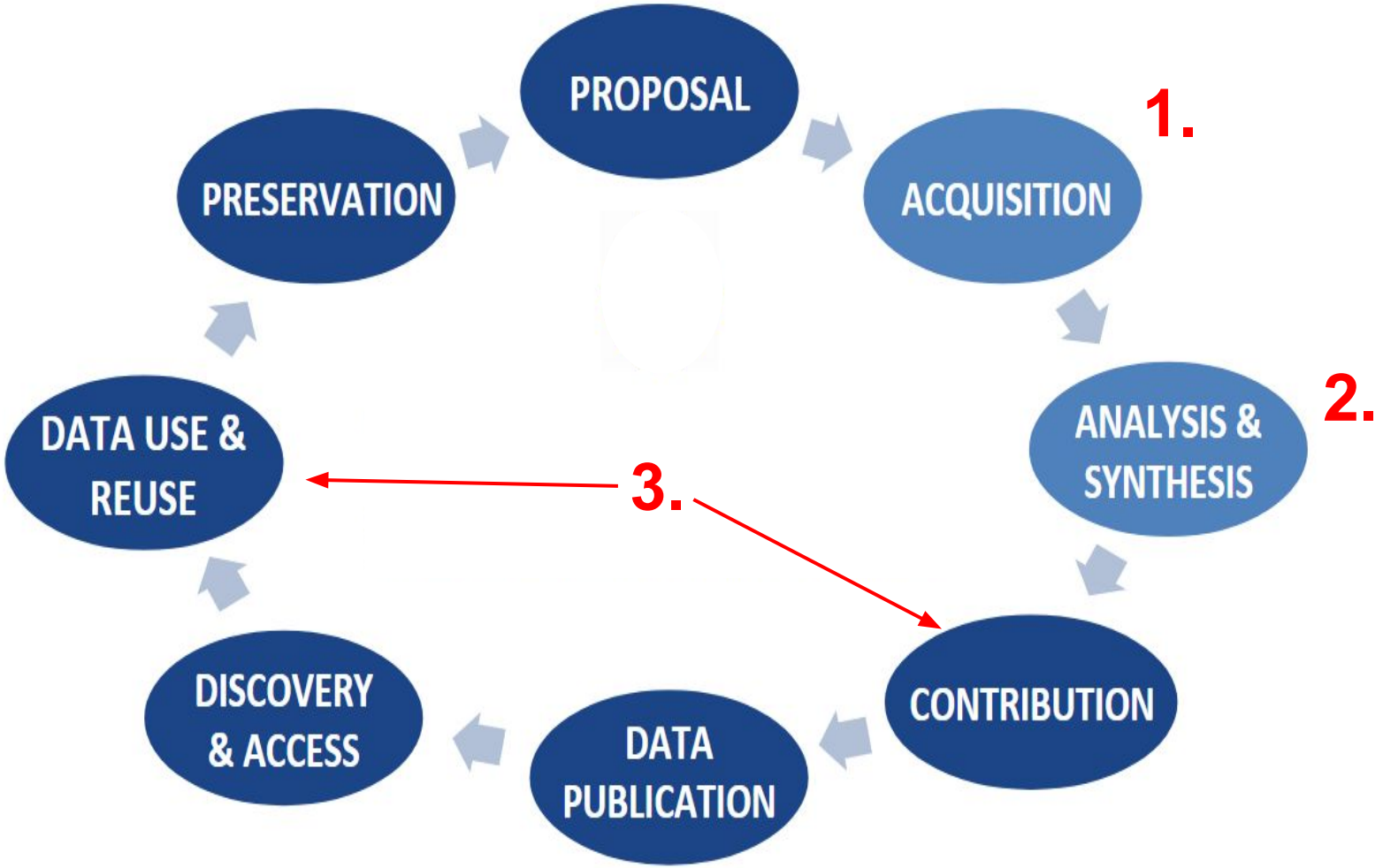
Navigating an Ocean of Data Training Camp

Module 4:
Best practices in the ocean sciences

Stace Beaulieu and Danie Kinkade

January 23, 2020

Module 4



THE DATA LIFE CYCLE

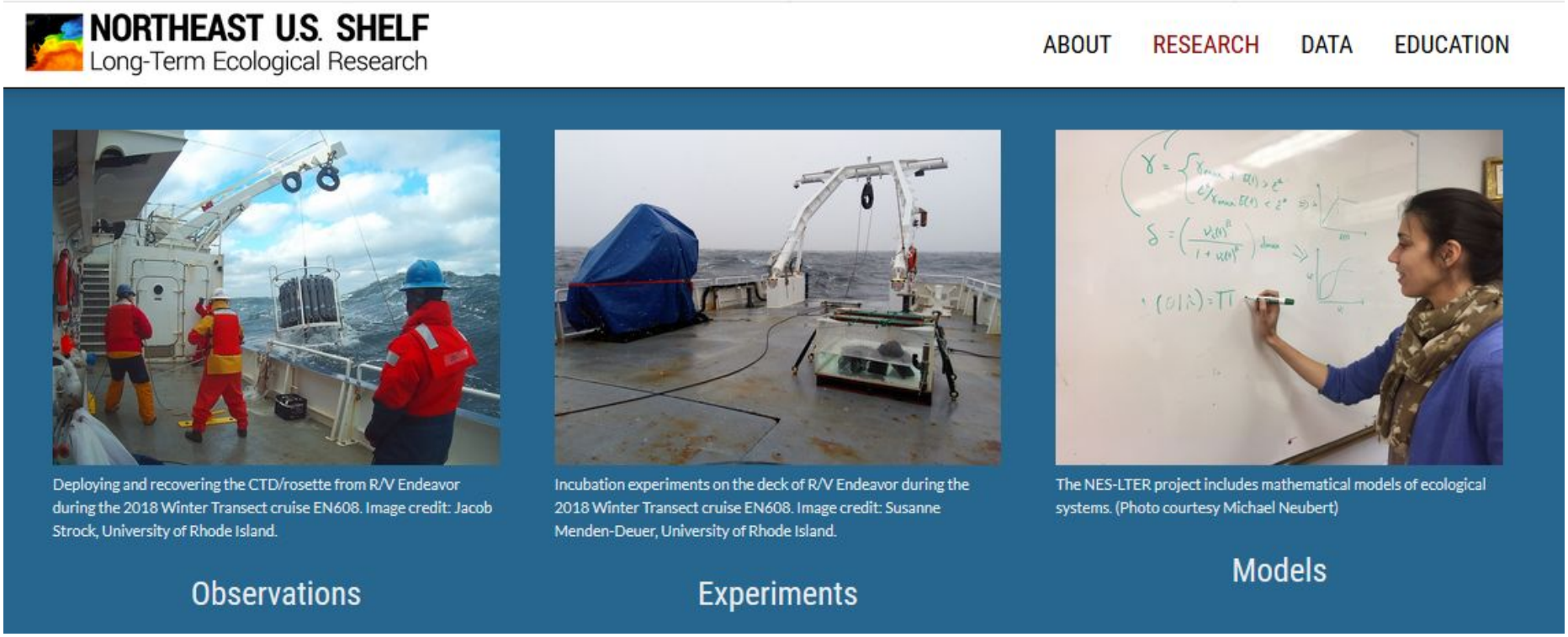
How are we different than the definition provided for “data scientist”?

- Data Scientists develop methods for storing, analyzing, and presenting data

Kempler, S., Mathews, T., (2017) Earth Science Data Analytics: Definitions, Techniques and Skills. *Data Science Journal*, 16(1), p.6, http://doi.org/10.5334/dsj-2017-006

We are data creators!

Types of data created at WHOI range from observational (sensors and physical samples) to experiments (with observations) to models



Example: NES-LTER project (https://nes-lter.whoi.edu/)

We are data creators!

http://www.whoi.edu/data/

Woods Hole Oceanographic Institution

PRESS ROOM | SHOP WHOI | DIRECTORY

WHO WE ARE | WHAT WE DO | KNOW YOUR OCEAN | JOIN US | DONATE

WHAT WE DO

- Understand
  - Areas of Research
  - Departments, Centers & Labs
  - Programs & Projects
  - Scientific Facilities & Services
  - Data & Repositories
    - Other National and International Repositories
- Explore

SHARE THIS:

Data & Repositories

Data and Repositories Based at WHOI

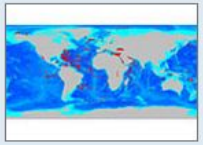
Alvin Frame-Grabber

Contacts: Steve Lerner

The Frame-Grabber imaging system mounted on the submersible Alvin provides Web access to video imagery co-registered with vehicle navigation and attitude data for shipboard analysis, planning deep-submergence research cruises, and review of data following research expeditions.

Argo Float Data

View the locations of WHOI's floats being used as part of the ARGO project to observe the ocean and to predict climate change. See their launch positions, track the paths they have followed, and pinpoint the locations of



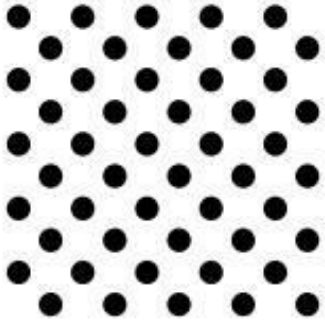
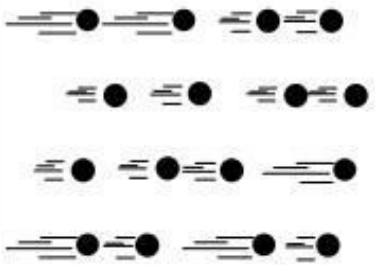
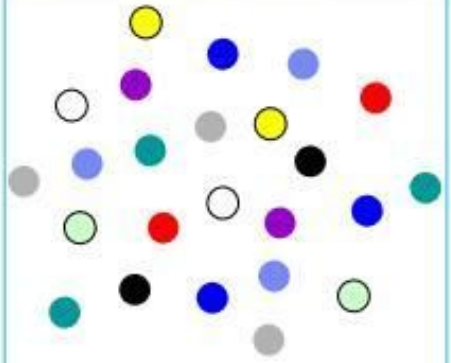
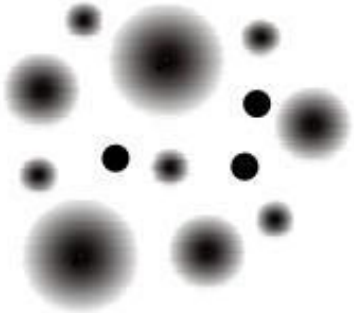
Seafloor Sediments Data Collection

The collection and analysis efforts of WHOI's Seafloor Sediments Laboratory are available for browsing. A collection of more than 14,000 archived marine geological samples recovered from the seafloor. The inventory includes long, stratified sediment cores, as well as rock dredges, surface grabs, and samples collected by the submersible Alvin.

Upper Ocean Processes Mooring Data

Data collected during the past 25 years by UOP surface moorings that measure meteorological conditions and physical properties of the upper ocean. Includes data gathered along the coasts and the open ocean during various weather conditions and from a range of geographic regions, from subarctic regions to the tropics.

You experience Big Data when your ability to manage the data exceeds any of these V's:

Volume	Velocity	Variety	Veracity*
			
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

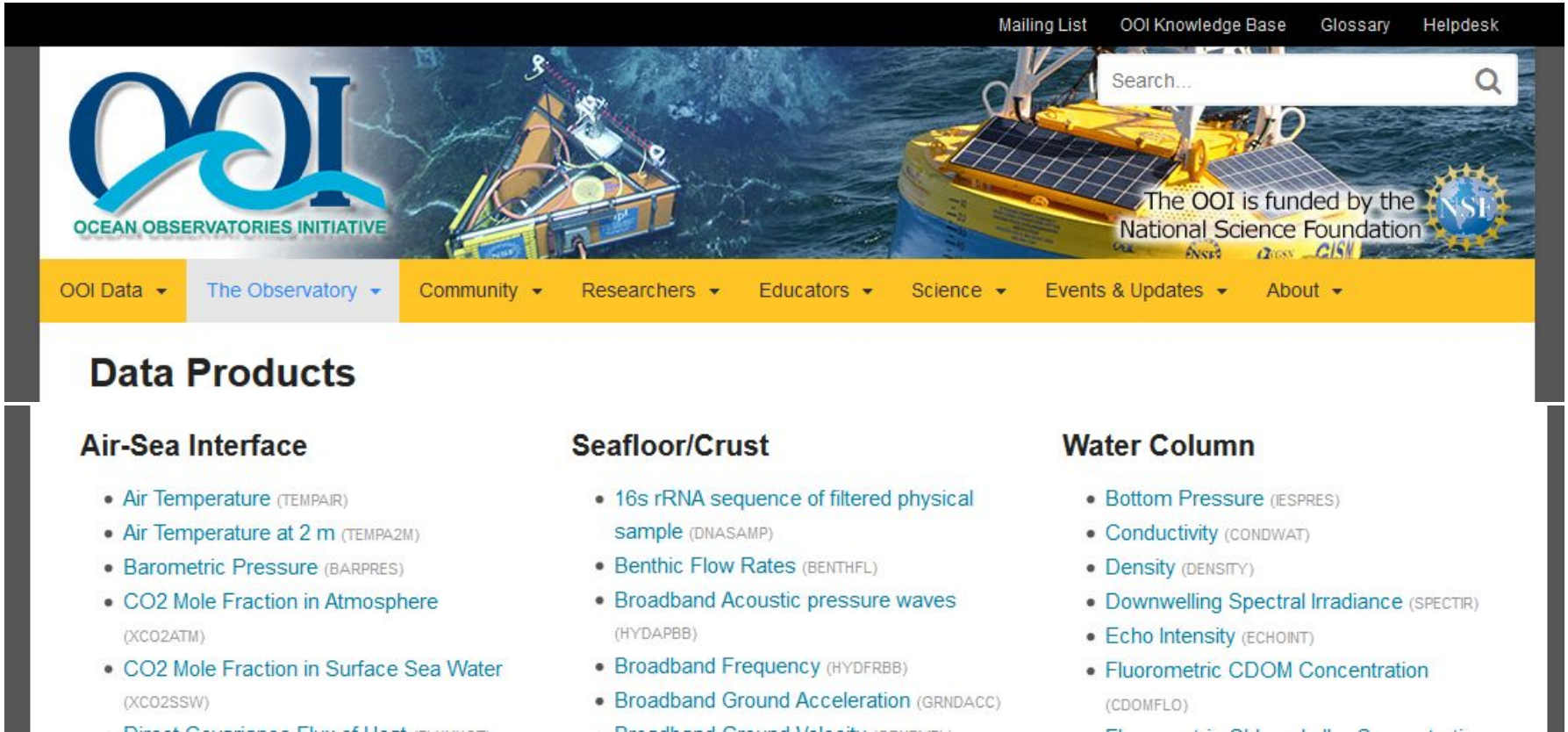
https://www.datasciencecentral.com/profiles/blogs/data-veracity

We create a variety of high volume and high velocity data

Module 3 highlighted high volume data

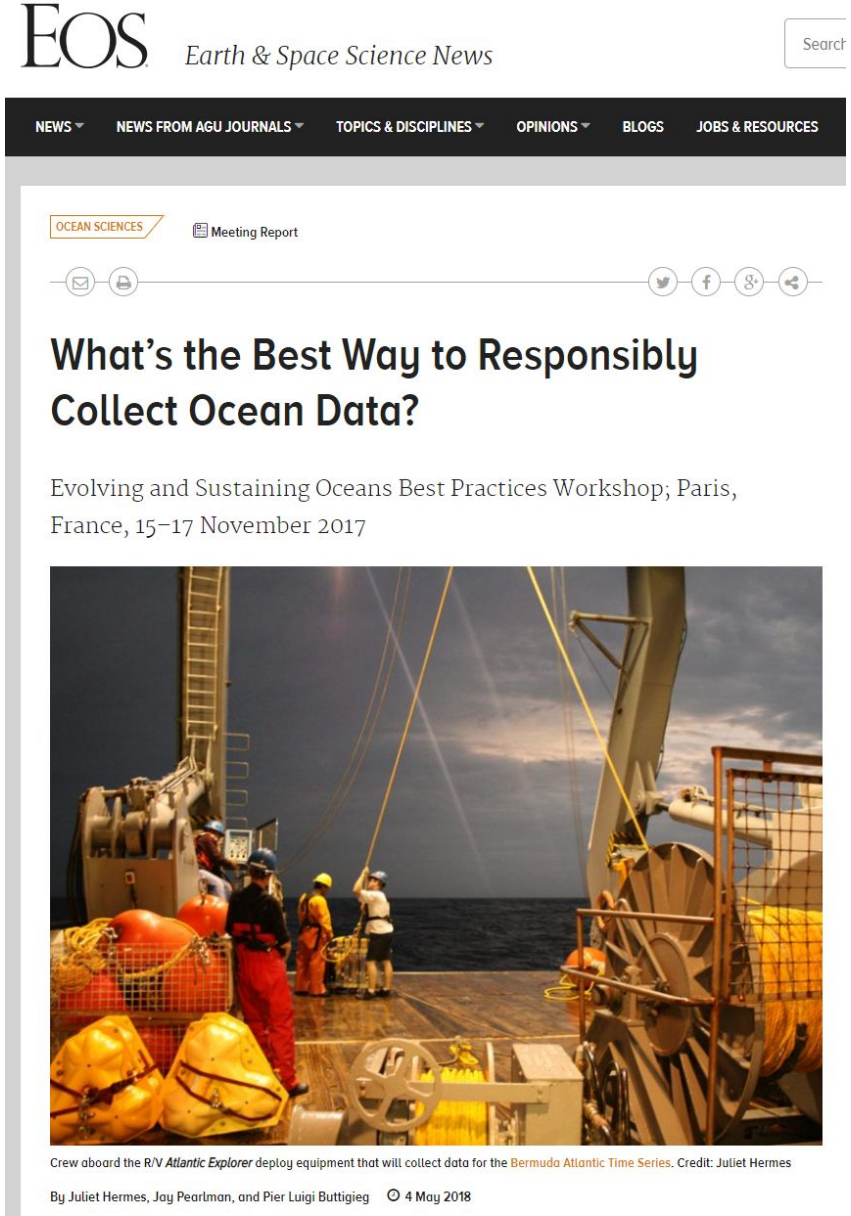
e.g., near-real-time observations

Example: https://oceanobservatories.org/data/



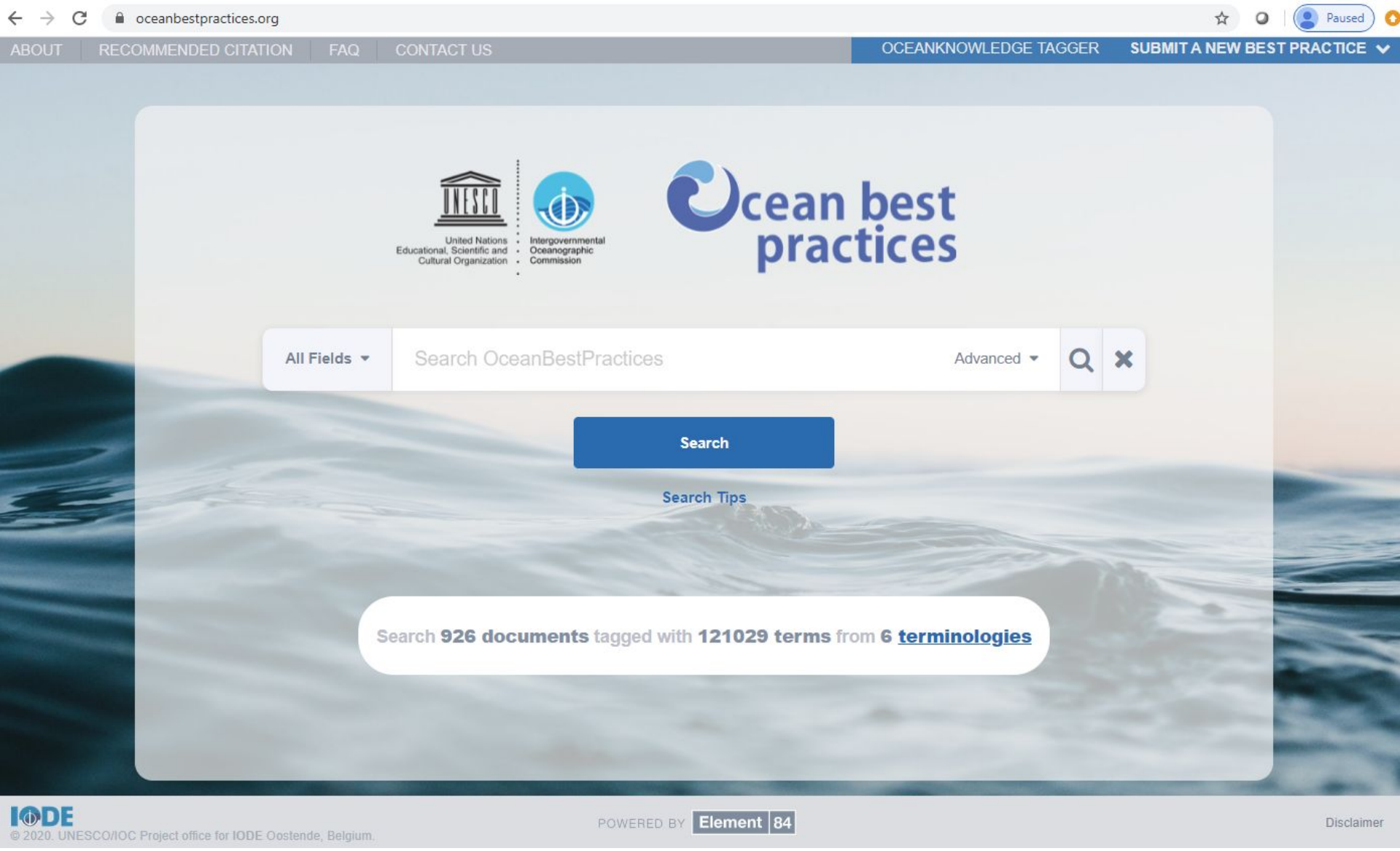
We have best practices for creating these data

"A community best practice is a methodology that has repeatedly produced superior results relative to other methodologies with the same objective."



https://doi.org/10.1029/2018EO096533

https://www.oceanbestpractices.org/



We may use physical samples to create these data

It's not just data that need an identifier…
it's also the physical samples from which the data derive!

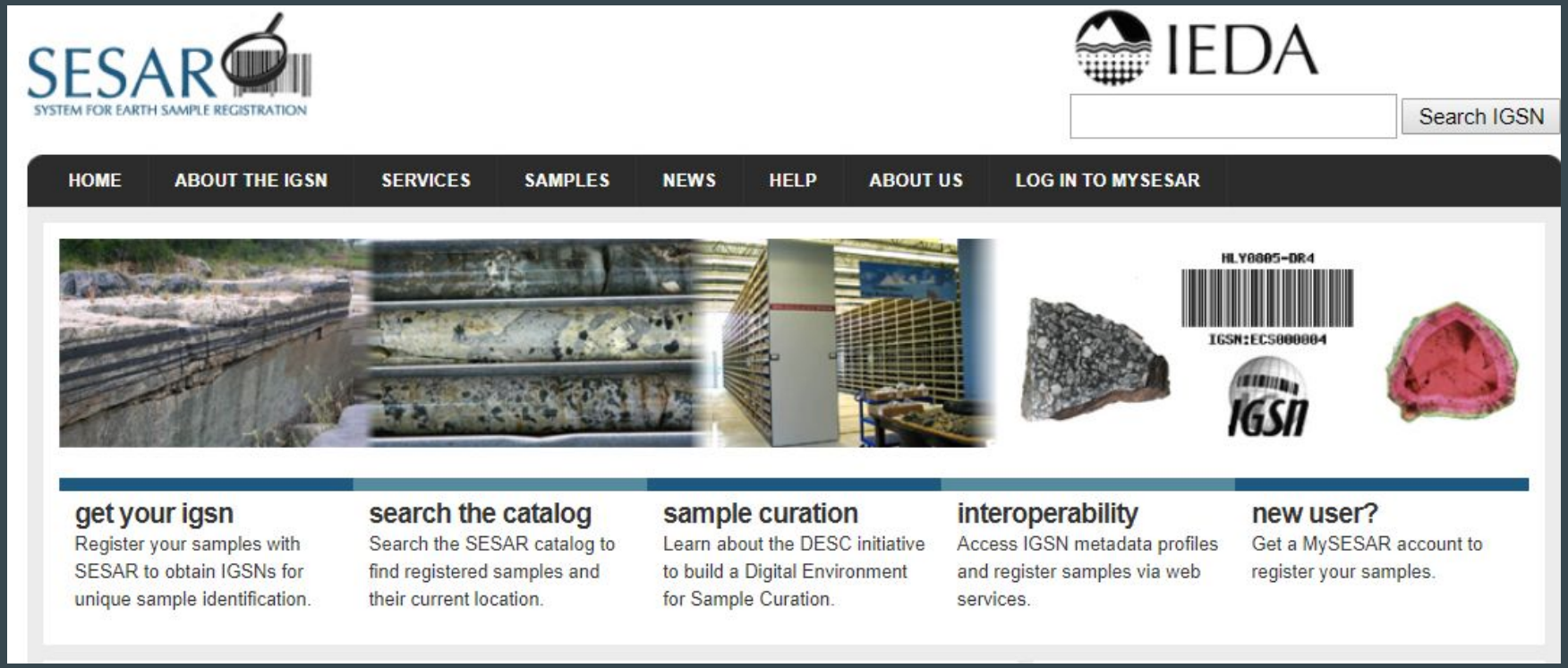
Photo by Meg Tivey

Email from Meg when we sent out last year's workshop announcement:
"Perfect timing (for me). I am filling out my IGSN metadata registration excel spreadsheets at this moment!"

International Geo Sample Numbers (IGSNs)

The IGSN is a persistent unique identifier for physical samples and specimens that eliminates the problems associated with the ambiguous naming of samples. In the U.S., you can obtain IGSNs using the System for Earth Sample Registration (SESAR) at IEDA Data Facility.

See https://igsn.github.io/overview/ and http://www.geosamples.org/



Let's look for Meg's recently registered International Geo Sample Numbers (IGSNs)

In U.S.: System for Earth Sample Registration (SESAR) at IEDA Data Facility

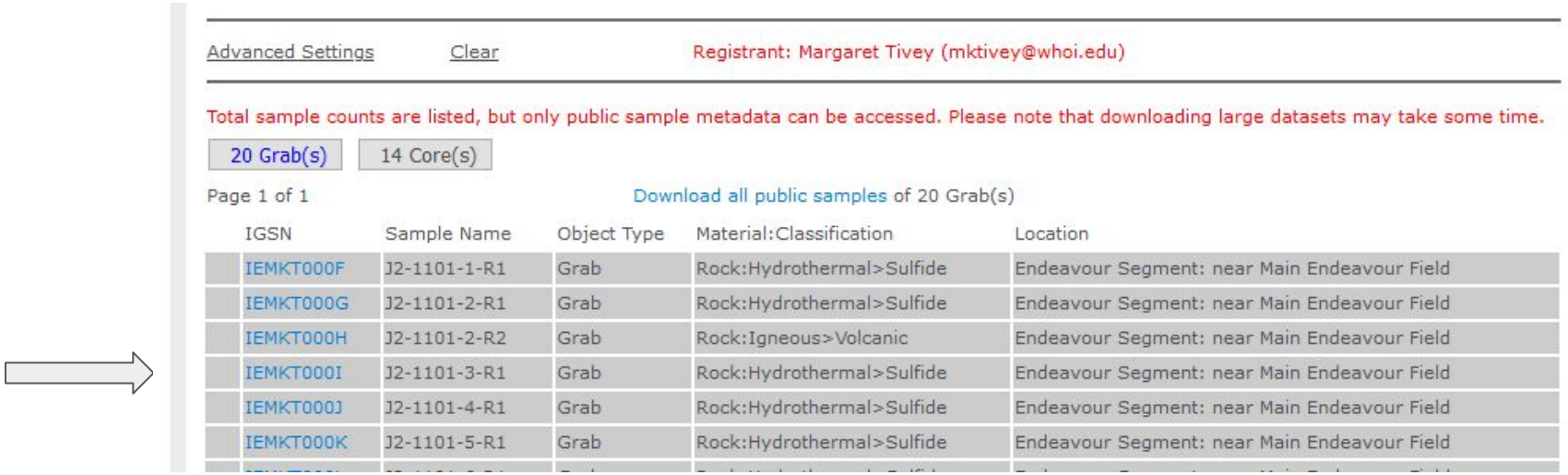
Take the next 5 min:

- Search for Meg's samples in SESAR: http://www.geosamples.org/catalogsearch
- What kinds of information can you find out about her samples?

Hint: Advanced Settings > Registrant

2:45pm

What kinds of information can you find out about her samples?



Advanced Settings Clear Registrant: Margaret Tivey (mktivey@whoi.edu)

Total sample counts are listed, but only public sample metadata can be accessed. Please note that downloading large datasets may take some time.

20 Grab(s) 14 Core(s)

Page 1 of 1 Download all public samples of 20 Grab(s)

IGSN	Sample Name	Object Type	Material:Classification	Location
IEMKT000F	J2-1101-1-R1	Grab	Rock:Hydrothermal>Sulfide	Endeavour Segment: near Main Endeavour Field
IEMKT000G	J2-1101-2-R1	Grab	Rock:Hydrothermal>Sulfide	Endeavour Segment: near Main Endeavour Field
IEMKT000H	J2-1101-2-R2	Grab	Rock:Igneous>Volcanic	Endeavour Segment: near Main Endeavour Field
IEMKT000I	J2-1101-3-R1	Grab	Rock:Hydrothermal>Sulfide	Endeavour Segment: near Main Endeavour Field
IEMKT000J	J2-1101-4-R1	Grab	Rock:Hydrothermal>Sulfide	Endeavour Segment: near Main Endeavour Field
IEMKT000K	J2-1101-5-R1	Grab	Rock:Hydrothermal>Sulfide	Endeavour Segment: near Main Endeavour Field

IGSN: IEMKT000I

IGSN: IEMKT000I
Sample Name: J2-1101-3-R1
Other Name(s):
Sample Type: Grab
Parent IGSN: Not Provided

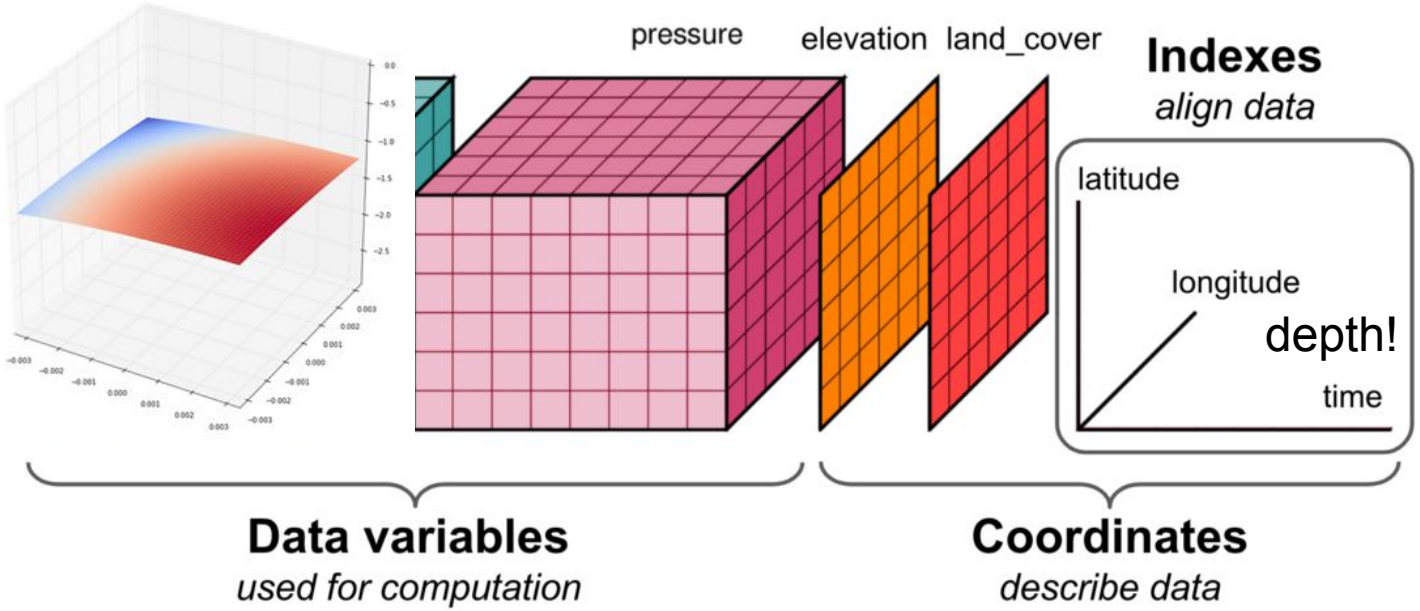
Description

Material: Rock

Latitude (WGS84): 47.9552676
Longitude (WGS84): -129.0960122
Elevation Start: -2160.73 meters
Field Program/Cruise: KM1812
Platform Type: Ship
Platform Name: Kilo Moana
Launch Type: ROV
Launch Platform Name: Jason II
Launch ID: J2-1101
Collection Start Date: Not Provided

Importance of the 4 D's in our domain

dimensions



From Module 1, definition Earth Science Data Analytics:
"The process of examining, preparing, reducing, and analyzing large amounts of spatial (multi-dimensional), temporal, or spectral data, encompassing varieties of data types..."

https://medium.com/pangeo/step-by-step-guide-to-building-a-big-data-portal-e262af1c2977
(blog post by MIT alum Ryan Abernathy)

Controlled vocabularies for (units and) variables

NERC Vocabulary Server

https://www.bodc.ac.uk/resources/vocabularies/

Poll results from yesterday:

Have you heard of a "controlled vocabulary"?

- Yes = 0?

F and I in FAIR data

e.g., http://vocab.nerc.ac.uk/collection/P02/current/SIGT/

⇑ -- Density of the water column --

URI	http://vocab.nerc.ac.uk/collection/P02/current/SIGT/
Identifier ()	SDN:P02::SIGT
Preferred label (en)	**Density of the water column**
Alternative label ()	WC_Dens
Version Info ()	1
Has Current Version	http://vocab.nerc.ac.uk/collection/P02/current/SIGT/1/
	1
)	2004-08-16 11:31:09.0
	Absolute determinations of water column density plus parameter
	false
	http://vocab.nerc.ac.uk/collection/P03/current/D020/
	http://vocab.nerc.ac.uk/collection/D01/current/D0100001/
	http://vocab.nerc.ac.uk/collection/L19/current/005/
	http://vocab.nerc.ac.uk/collection/L04/current/L04001/
	http://vocab.nerc.ac.uk/collection/P05/current/014/
	http://vocab.nerc.ac.uk/collection/P22/current/28/
	http://vocab.nerc.ac.uk/collection/P01/current/SIGTSG01/
	http://vocab.nerc.ac.uk/collection/P07/current/BBAH2104/

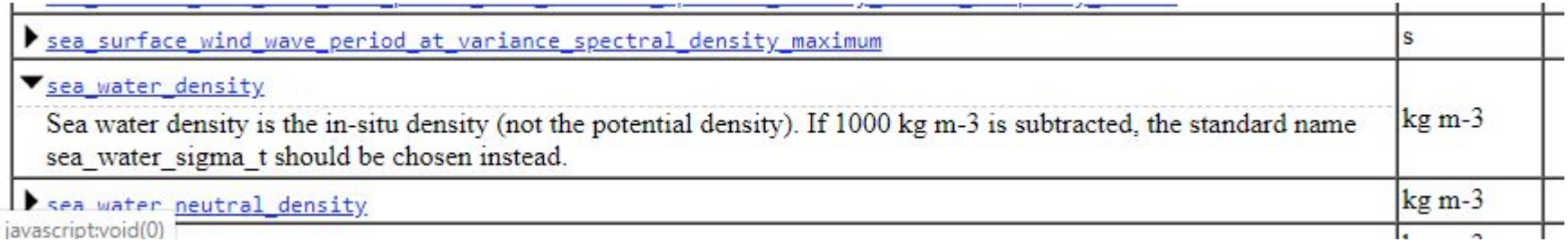
⇑ -- Sigma-theta of the water body by thermosalinograph and computation from salinity and potential temperature using UNESCO algorithm --

URI	http://vocab.nerc.ac.uk/collection/P01/current/SIGTSG01/
Identifier ()	SDN:P01::SIGTSG01
Preferred label (en)	**Sigma-theta of the water body by thermosalinograph and computation from salinity and potential temperature using UNESCO algorithm**
Alternative label (en)	STTSG
Definition (en)	Computed by UNESCO SVAN function using potential temperature
Version Info ()	1
Has Current Version	http://vocab.nerc.ac.uk/collection/P01/current/SIGTSG01/1/
PAV Version ()	1
PAV Authored On ()	2009-11-03 16:19:38.0
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/P02/current/SIGT/
Broader	http://vocab.nerc.ac.uk/collection/S26/current/MAT00640/

Controlled vocabularies in our domain

e.g., Climate and forecast (CF) metadata convention

http://cfconventions.org/standard-names.html



sea_surface_wind_wave_period_at_variance_spectral_density_maximum	s
sea_water_density Sea water density is the in-situ density (not the potential density). If 1000 kg m-3 is subtracted, the standard name sea_water_sigma_t should be chosen instead.	kg m-3
sea_water_neutral_density	kg m-3

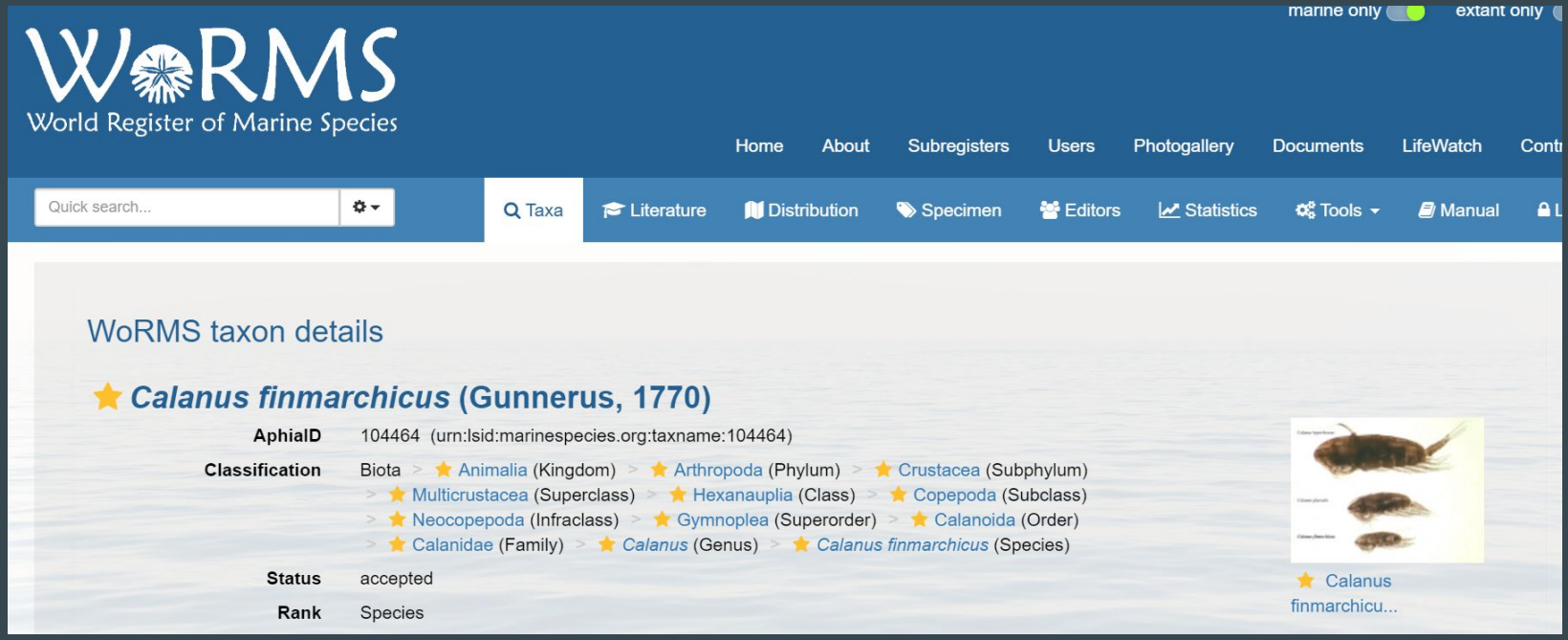
javascript:void(0)

Other examples:

- acoustic_signal_roundtrip_travel_time_in_sea_water
- mass_concentration_of_nitrous_oxide_in_air

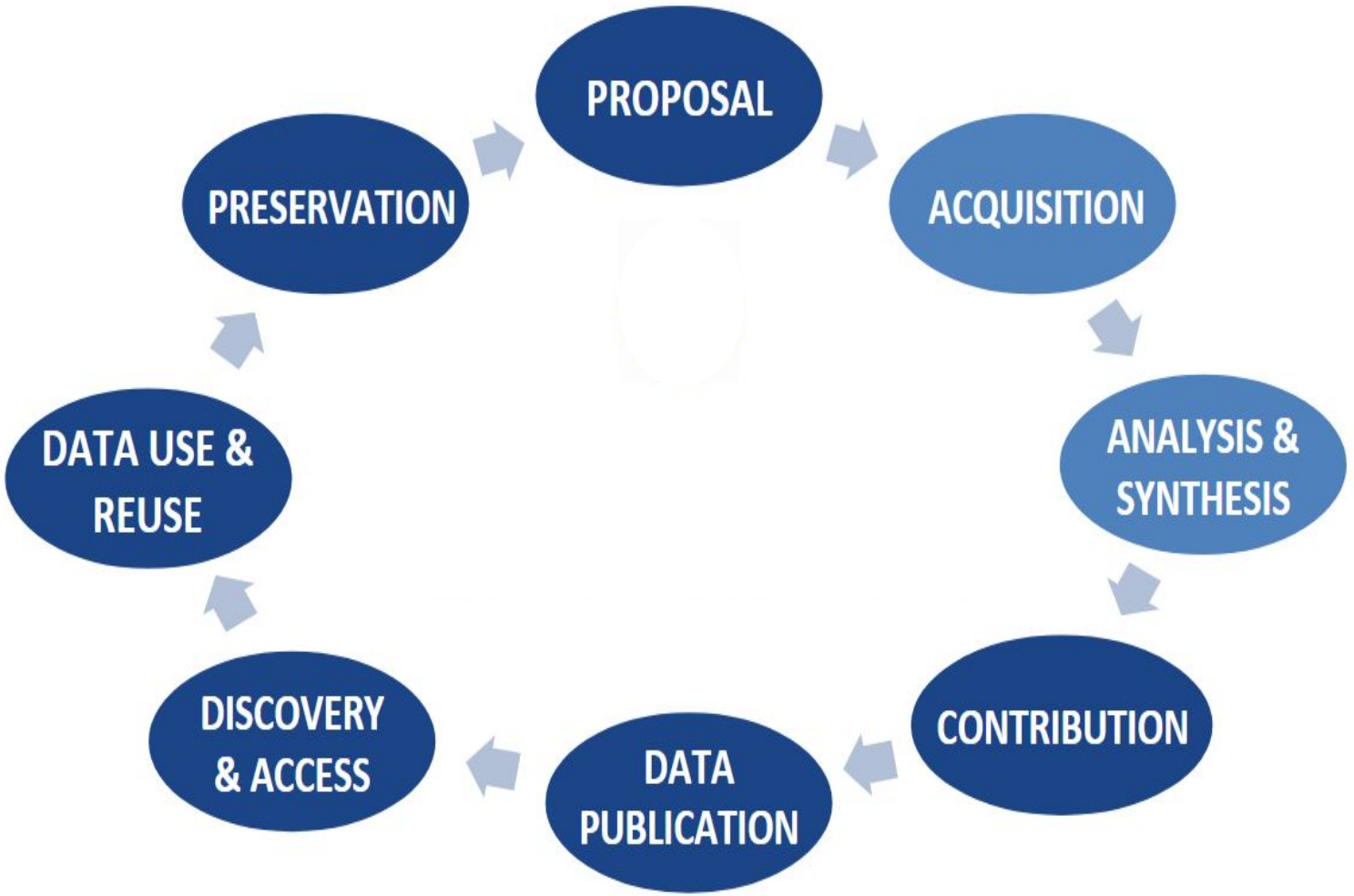
Taxonomy - WoRMs

The World Register of Marine Species (WoRMs) provides an authoritative and comprehensive list of names of marine organisms, including information on synonymy.



We recommend checking species names in WoRMs and including identifiers in your data when possible.

Module 4

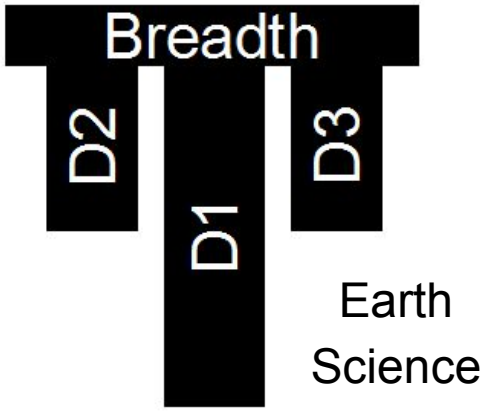


2.

THE DATA LIFE CYCLE

We store, analyze, and present data… what else makes us different than data scientists outside our domain?

Data, statistics & computing skills

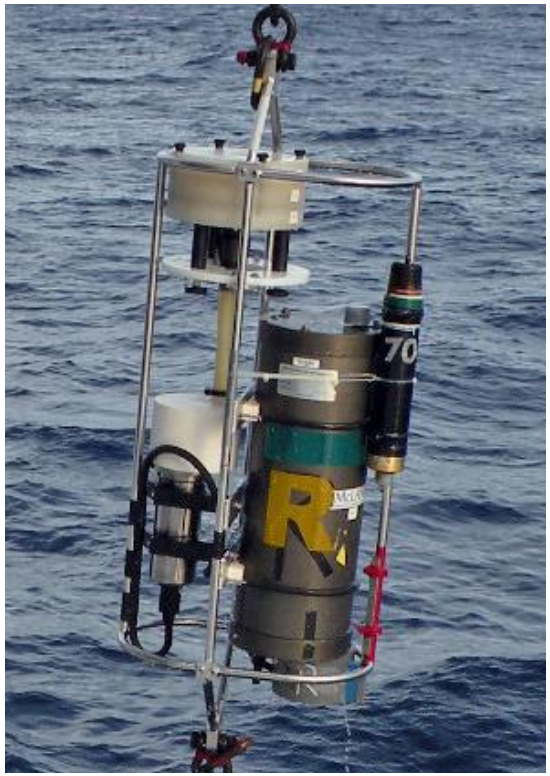


shield-shaped
1 person

Deep knowledge of how to appropriately use the data

We have knowledge and/or intuition about data quality and data uncertainty

To distinguish these concepts, watch Carol Anne Clayson's plenary talk "Information about the data is as important as the data itself" from 2017 ESIP Summer Meeting, 3:22 to 3:34 in: https://www.youtube.com/watch?v=8fP4M0iAYGs



We get to know our instruments, e.g., "Ringo"

Quality Flags

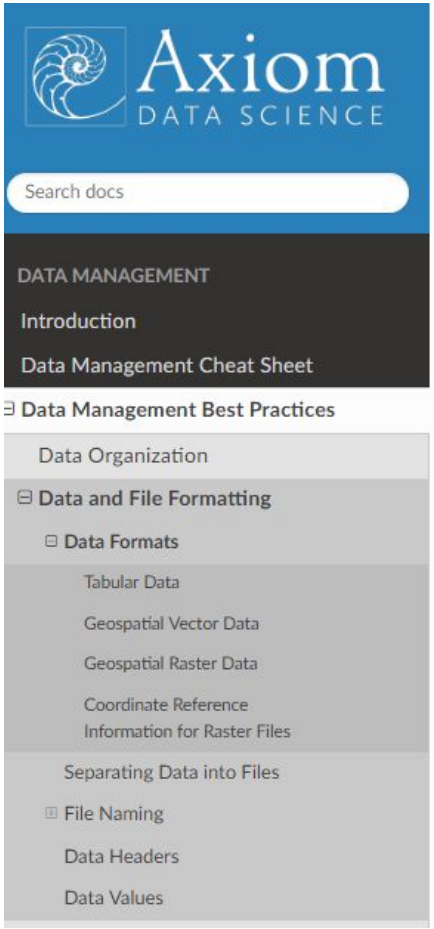
You can incorporate data quality and uncertainty into your data and/or metadata.

e.g., International Oceanographic Data and Information Exchange (IODE) quality flags.
See: https://www.iode.org/mg54_3

Value	Primary-level flag short name	Definition
1	Good	Passed documented required QC tests
2	Not evaluated, not available or unknown	Used for data when no QC test performed or the information on quality is not available
3	Questionable/suspect	Failed non-critical documented metric or subjective test(s)
4	Bad	Failed critical documented QC test(s) or as assigned by the data provider
9	Missing data	Used as place holder when data are missing

We encounter a variety of data formats in our domain

https://www.axiomdatascience.com/best-practices/DataandFileFormatting.html#data-formats



Data and File Formatting

Data Formats

In choosing a file format, data collectors should select a format that is useable, open, and that will likely be readable well into the future. Microsoft Excel, as an example, is a useful tool for data manipulations and data visualization, but versions of Excel files may become obsolete and may not be easily readable over the longer term. Likewise, database management systems (DBMS) like MS Access, Filemaker Pro, and others, can be a very effective way to store and query data, but the raw formats tend to change over time (even a few years). If your program or organization has used these or other proprietary DBMS tools, it is essential to plan for exporting your data in a stable, well-documented, and non-proprietary format.

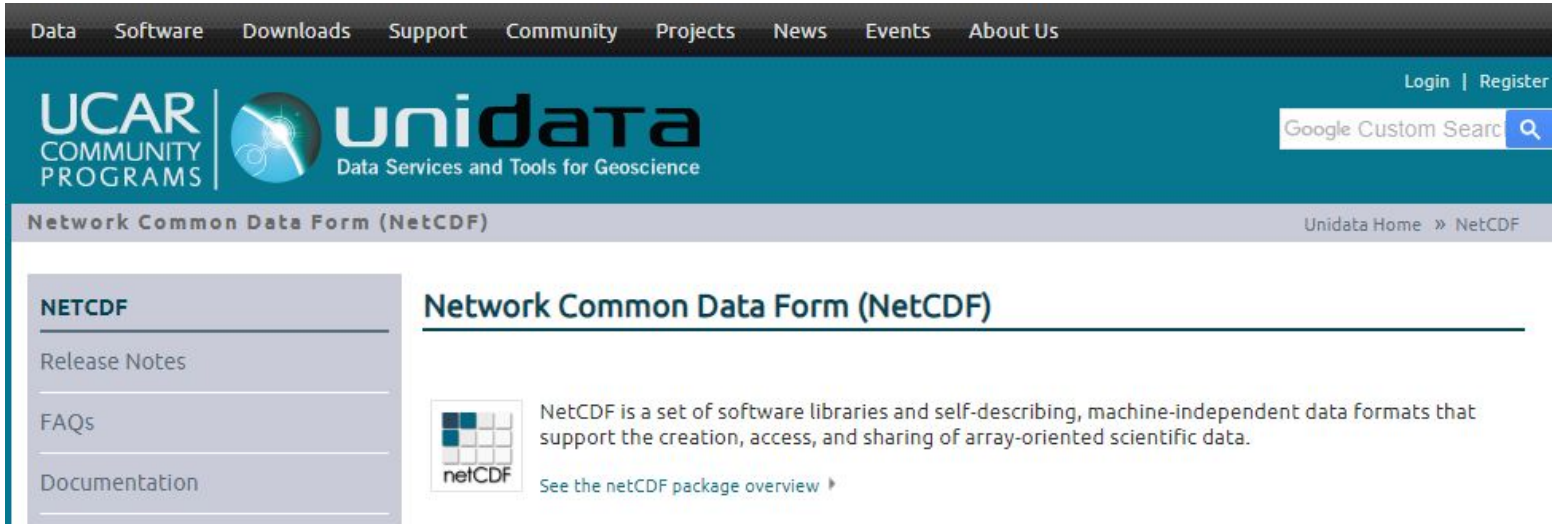
Below is a summary of the suggested tabular, image, and GIS data file formats suitable for long-term archiving.

- Containers: TAR, GZIP, ZIP
- Databases: CSV, XML
- Tabular data: CSV
- Geospatial vector data: SHP, GeoJSON, KML, DBF, NetCDF
- Geospatial raster data: GeoTIFF/TIFF, NetCDF, HDF-EOS
- Moving images: MOV, MPEG, AVI, MXF
- Sounds: WAVE, AIFF, MP3, MXF
- Statistics: ASCII, DTA, POR, SAS, SAV
- Still images: TIFF, JPEG 2000, PDF, PNG, GIF, BMP
- Text: XML, PDF/A, HTML, ASCII, UTF-8
- Web archive: WARC

We are highlighting CSV and NetCDF in this workshop

Data formats in our domain

e.g., NetCDF (https://www.unidata.ucar.edu/software/netcdf/)



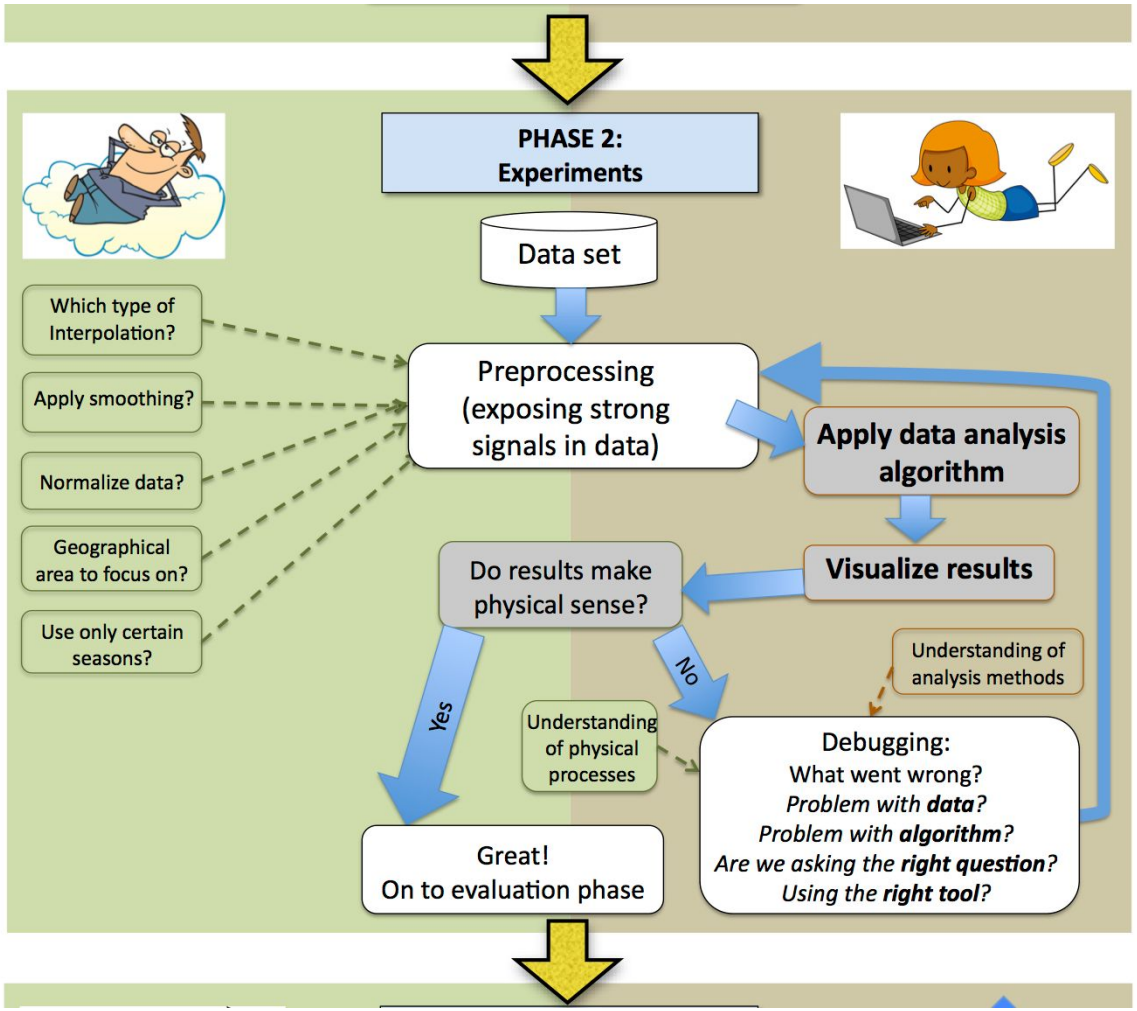
NetCDF (.nc) is a standard delivery format for OOI, PO.DAAC, BODC, and other repos in our domain

NCEI templates: https://www.nodc.noaa.gov/data/formats/netcdf/v2.0/

The conventions for CF (Climate and Forecast) metadata are designed to promote the processing and sharing of files created with the NetCDF API.
http://cfconventions.org/Data/cf-documents/overview/viewgraphs.pdf

"The ability to store and exchange data in a standardized way is an essential requirement to build the data processing e-infrastructure needed for Big Data analysis."

EU Marine Board
Future Science Brief



Ebert-Uphoff, I., and Y. Deng (2017), Three steps to successful collaboration with data scientists, *Eos, 98,* https://doi.org/10.1029/2017EO079977

Types of Analyses

- Processing: Subsetting, Merging, Manipulating
- Statistical analyses
- Graphical analyses and visualization

Data Analysis and Workflows:
https://dataoneorg.github.io/Education/lessons/09_analysis/slides.html#1

Earth Science Data Analytics: "The process of examining, preparing, reducing, and analyzing large amounts of spatial (multi-dimensional), temporal, or spectral data, encompassing varieties of data types... to uncover patterns, correlations, and other information, to better understand our Earth."

Kempler, S., Mathews, T., (2017) Earth Science Data Analytics: Definitions, Techniques and Skills. *Data Science Journal*, 16(1), p.6, http://doi.org/10.5334/dsj-2017-006

When analyzing our own data, we often want to compare to data in other studies.

Were you able to find/access the data reported in your homework paper?

Elsevier includes protocols, code, and software in their definition of Research Data:
https://www.elsevier.com/authors/author-services/research-data

We encounter a variety of types of software to analyze data in our domain

e.g.,

ArcGIS

We use and create software to process/analyze/visualize data in our domain

How do you find out about new software that you might want to use?

Results of poll at the 2018 AGU Fall Meeting Town Hall for sharing scientific software:
people *either* found software in a Google search *or* they found out within their domain * (asked their colleagues, learned in papers, saw at conferences, etc.)

* this is why we have tables-by-department today for our discussion

Were you able to find/access the software used in your homework paper?

Take the next 10 min at your table to discuss with your peers the types of analyses in your homework paper, and whether you could access the software for those analyses

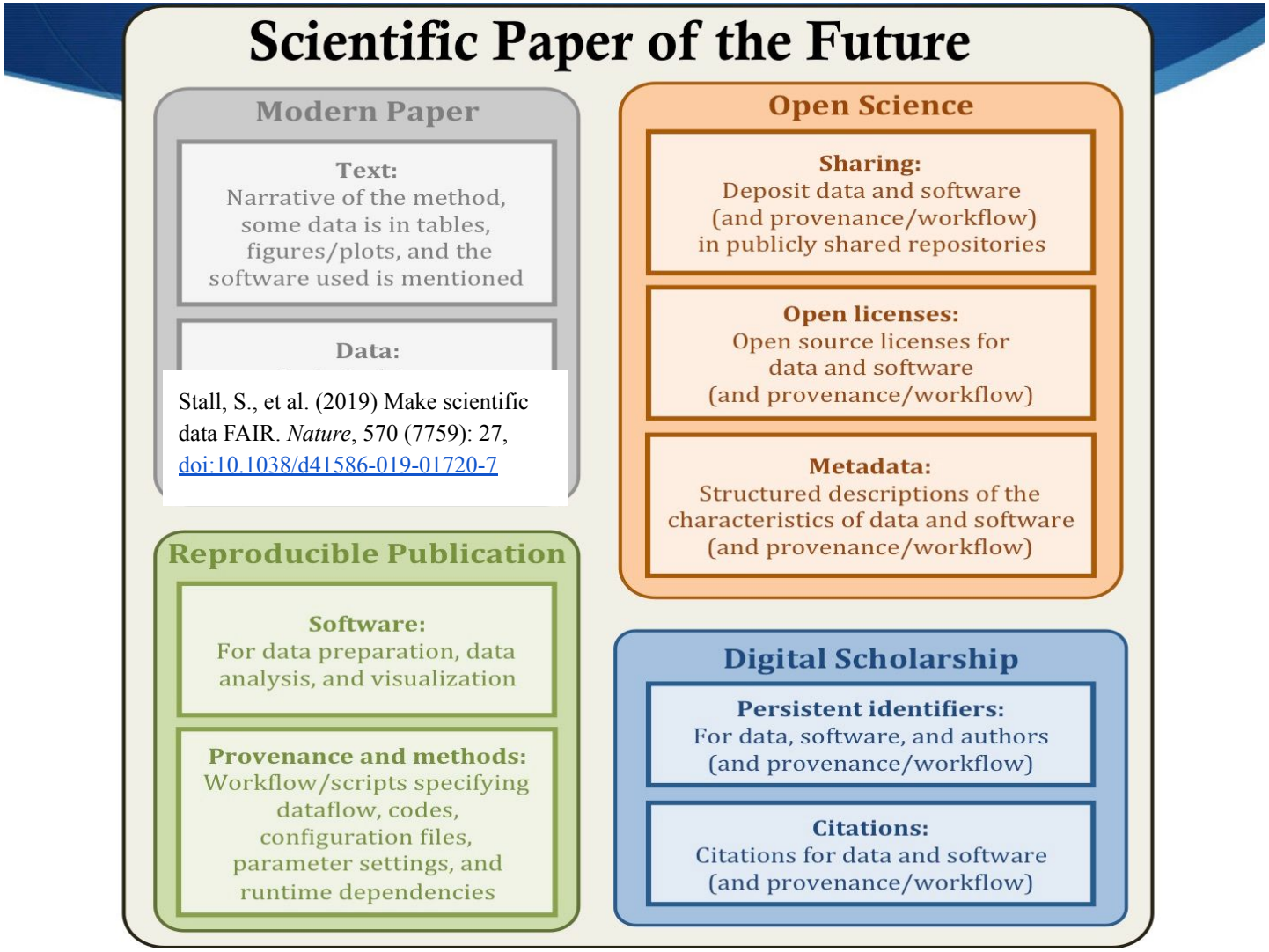
Add the journal for your homework paper to our shared notes

"welcoming and supportive environment for all"

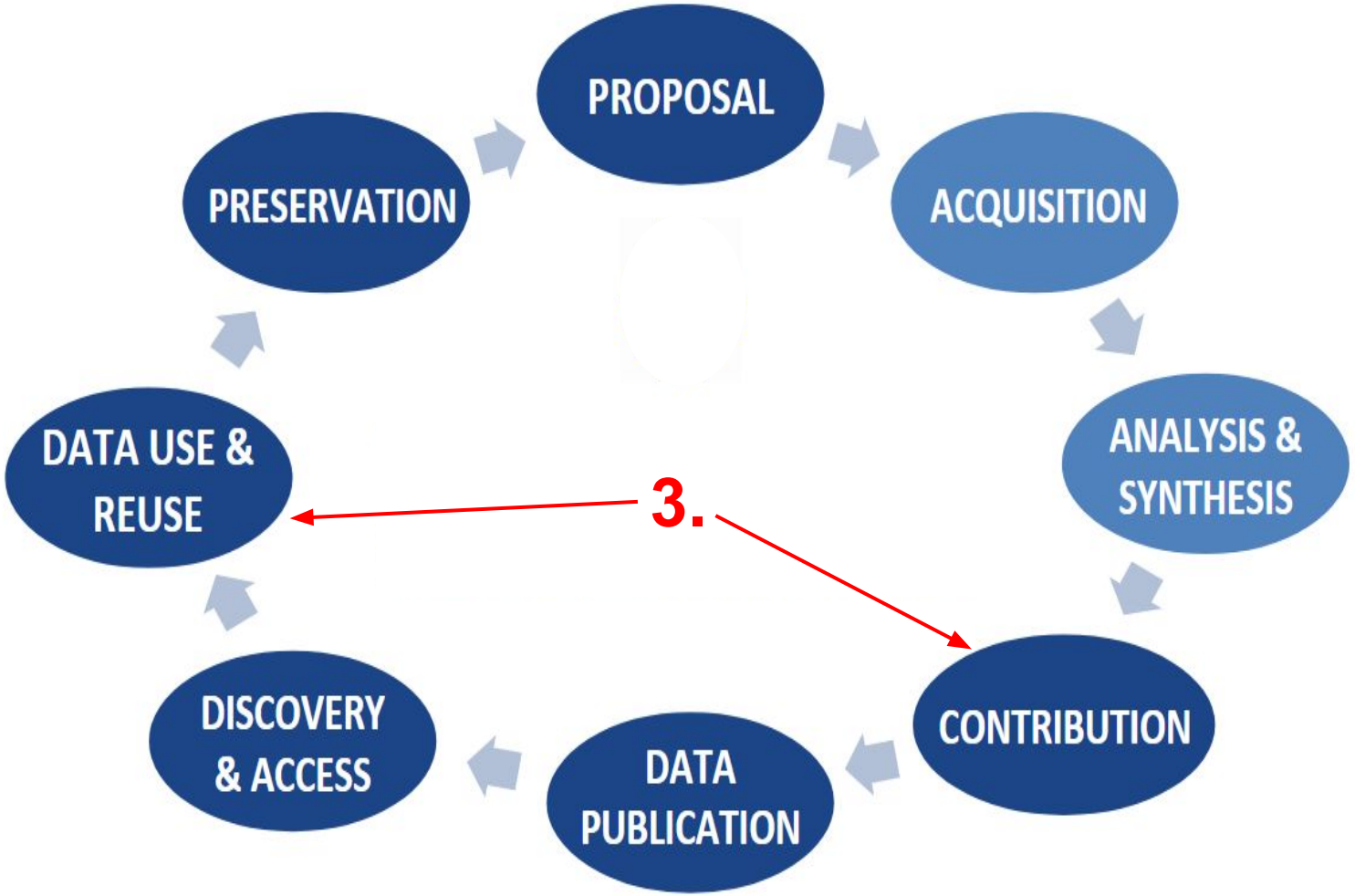
3:25pm Stretch!

Follow-up to homework

http://scientificpaperofthefuture.org/gpf/



Module 4



THE DATA LIFE CYCLE

3:40pm

“A scholar’s positive contribution is measured by the sum of the original data that [s/he] contributes. Hypotheses come and go but data remain.”

from: Advice to a Young Investigator (Santiago Ramón y Cajal, 1897)

photo by Chris Linder (WHOI)

Results from yesterday's poll: Repositories that you have contributed to

NCEI:
National Geophysical Data Center (NGDC) and
Ocean Carbon Data System (OCADS)

Arctic Data Center (was ACADIS)

EDI

IRMA (National Parks Service)

Zenodo

How easy or difficult was the process of contributing data to a repository?

- Very easy
- Relatively easy
- Neutral
- Relatively difficult
- Very difficult

Why share your data?
Benefit to…

YOU:

- Credit for your hard work!
- Can increase citations
- Can boost collaborations
- Increases exposure
- Satisfies funder requirements

Research Community:

- Builds a community resource
- Enables new discoveries
- Sparks collaborations
- Allows for transparency and reproducibility of research results

Society at-large:

- Transparency boosts public confidence in scientific process
- Can contribute to management and policy
- Availability to audiences outside of research (education, general public)

7 HABITS FOR SHARING YOUR DATA WITH REPOSITORIES

1. Start Early!
2. Communicate with Collaborators
3. Connect with Funders
4. Seek out Available Resources
5. Reach out to Repo
6. Document in a DMP
7. Follow Through! (with data best practices)

FUNDER -- REPO -- RESEARCHER

National Science Foundation
WHERE DISCOVERIES BEGIN

Geosciences (GEO)

Email Print Share

OCE Policies on Data Management, Ship Time, Resubmission and Environmental Compliance

June 26, 2017

1. **NSF Proposal & Award Policies & Procedures Guide (PAPPG) including Guidance on Letters of Collaboration for Unfunded Collaborations**

2. **Data Management**

   - NSF Dissemination and Sharing of Research Results
   - Division of Ocean Sciences (OCE) Sample and Data Policy including preferred data and physical collection archives and centers
   - Biological and Chemical Oceanography Data Management Office (BCO-DMO): BCO-DMO has developed a Data Management Plan template to assist investigators in submission of plans that meet the NSF OCE Sample and Data Policy requirements.
   - Interdisciplinary Earth Data Alliance (IEDA): IEDA Data Management Plan (DMP) Tool provides an easy way to generate a data management plan for inclusion in NSF proposals.

3. **Proposals That Include Ship Time**

   - OCE Guidance for Proposals that Include Ship Time
   - UNOLS Cruise Planning Information for Scientists

LOST?

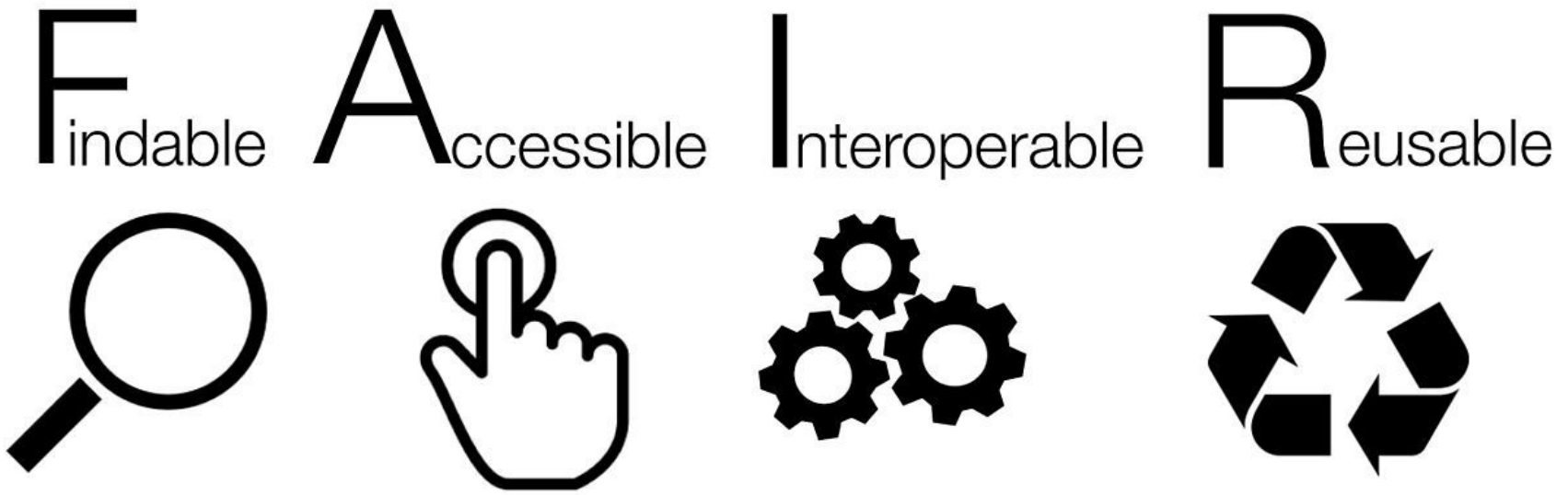
- What to do in the Absence of Guidance?

HABITS…

1. Domain
2. Institutional
3. Generic

Module 2: Finding Repositories (re3data)
https://www.re3data.org/

Recall the goal to make your data FAIR...

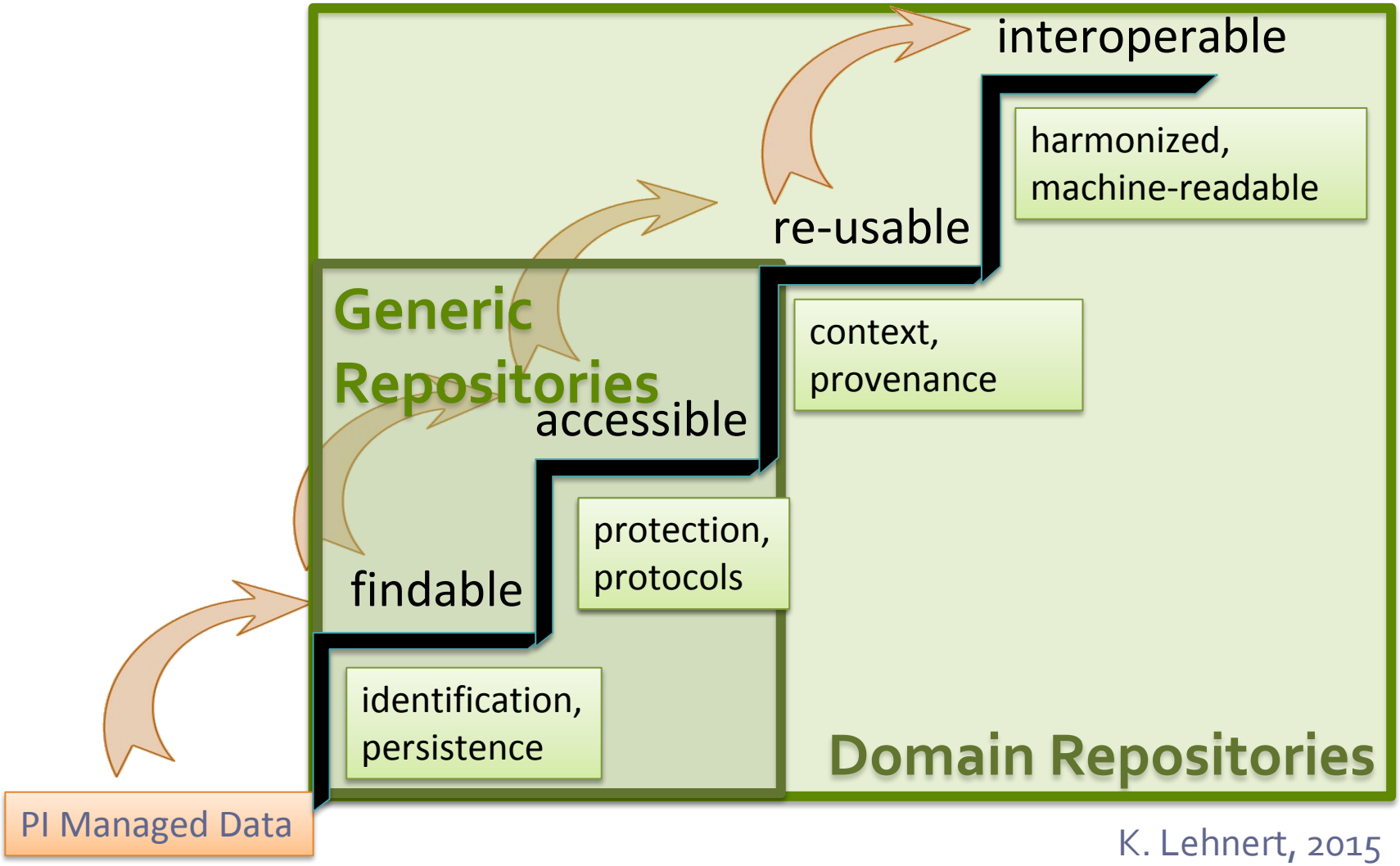


Original article describing FAIR data principles: Wilkinson, M.D., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, https://doi.org/10.1038/sdata.2016.18

Article more specific to FAIR in our domain: Stall, S., et al. (2018), Advancing FAIR data in Earth, space, and environmental science, *Eos*, 99, https://doi.org/10.1029/2018EO109301

HABITS…

FAIR Data



K. Lehnert, 2015

Activity for the “A”, “I”, and “R” in FAIR

What is ERDDAP? (simple)

- ERDDAP is a NOAA developed **data server**
- middle-man that facilitates **open** and **accessible** data download in a variety of formats applicable to various communities.
- **Accepts a variety of file formats** on the backend (ASCII text, netCDF, compressed files, image files, audio files, and others)
- Allows for **data and metadata** coupling and download
- Provides **web accessible urls** for advanced data download and subsetting via other software packages

Slide from “BCO-DMO's migration to ERDDAP,” DOI:10.1575/1912/24933, https://hdl.handle.net/1912/24933

“ERDDAP used to be an acronym, but it outgrew that original description. Now, please just think of it as a name, not an acronym.” https://coastwatch.pfeg.noaa.gov/erddap/information.html

Let's pull some data from BCO-DMO's ERDDAP directly into our software environment

Search BCO-DMO's ERDDAP for Dennis McGillicuddy (AOPE Chair)
https://erddap.bco-dmo.org/erddap/index.html

e.g.,
https://erddap.bco-dmo.org/erddap/tabledap/bcodmo_dataset_3358.html

BCO-DMO
Biological & Chemical Oceanography Data Management Office

BCO-DMO ERDDAP
Accessing BCO-DMO data

ERDDAP > search

Do a Full Text Search for Datasets:

dennis mcgillicuddy Search

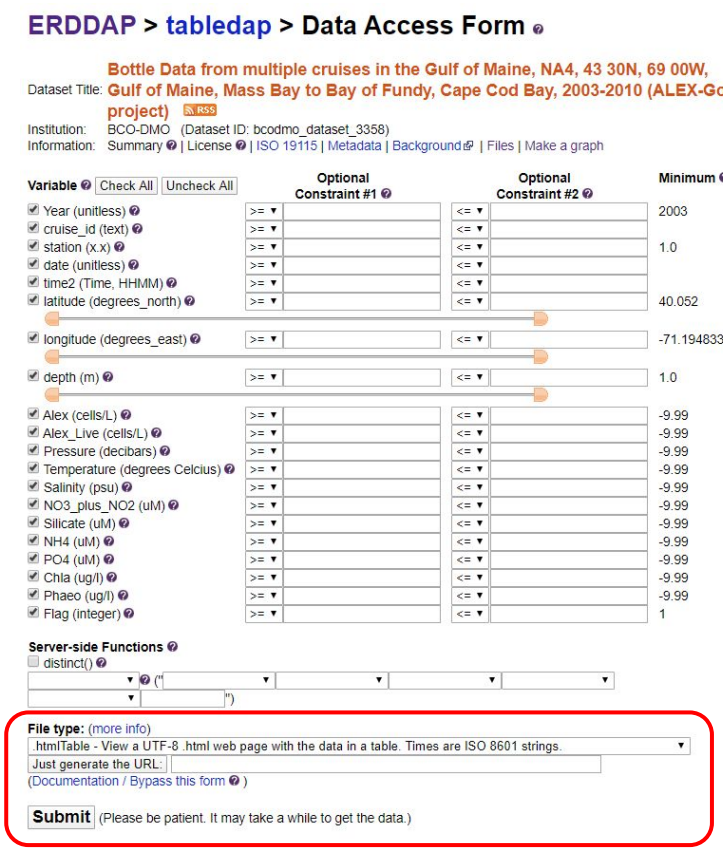
2 matching datasets, with the most relevant ones listed first.
(Or, refine this search with Advanced Search)

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	WMS	Source Data Files	Acces-sible	Title	Sum-mary	ISO, Metadata	Back-ground Info	RSS	Institution	Dataset ID
		data	graph		files	public	Pigments from HPLC analysis of bottle samples collected aboard Oceanus from R/V Oceanus OC404-01, OC404-04, OC415-01, OC415-03 in the Sargasso Sea from 2004-2005 (EDDIES project)		I M	background	RSS	BCO-DMO	bcodmo_dataset_3024
		data	graph		files	public	Bottle Data from multiple cruises in the Gulf of Maine, NA4, 43 30N, 69 00W, Gulf of Maine, Mass Bay to Bay of Fundy, Cape Cod Bay, 2003-2010 (ALEX-GoME project)		I M	background	RSS	BCO-DMO	bcodmo_dataset_3358

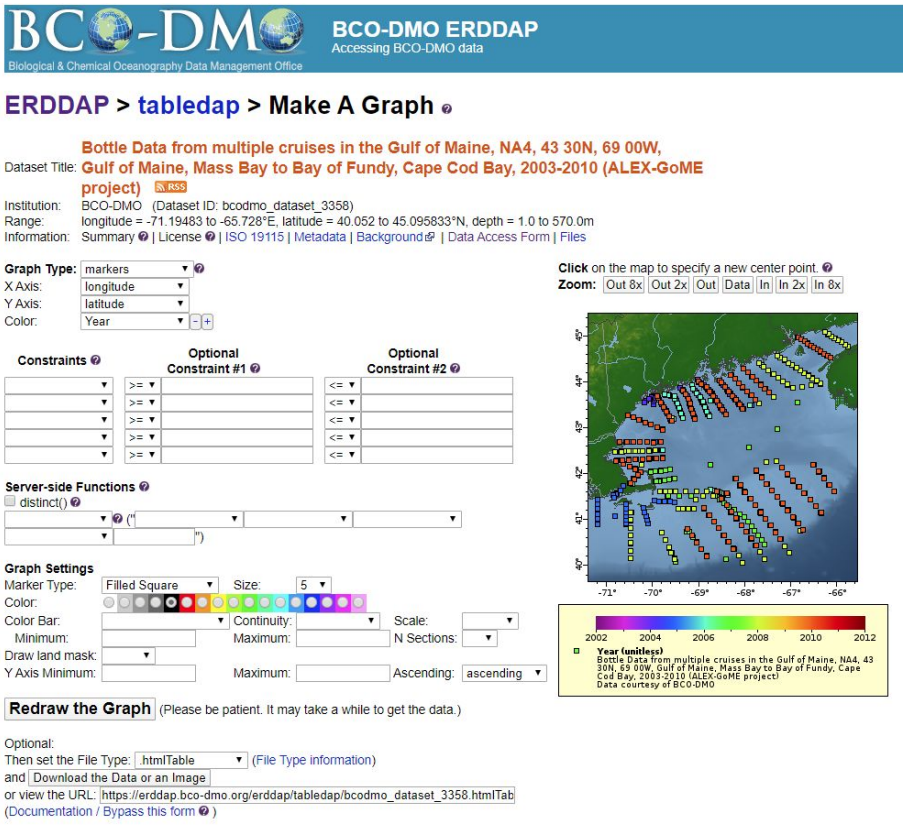
Let's pull some data from BCO-DMO's ERDDAP

Click on "data"
to generate URL

or graph
to explore in ERDDAP



Choose csvp (or .nc)



https://erddap.bco-dmo.org/erddap/tabledap/bcodmo_dataset_3358.html

Take the next 10 min to access data via Application Programming Interface (API)

ERDDAP URL pattern
https://*namespace*/erddap/tabledap/*datasetID*.*fileType*{?*query*}

Example URL:
https://erddap.bco-dmo.org/erddap/tabledap/bcodmo_dataset_3358.csvp?Year%2Ccruise_id%2Cstation%2Cdate%2Ctime2%2Clatitude%2Clongitude%2Cdepth%2CAlex_Live%2CTemperature%2CNO3_plus_NO2%2CChla

```
DennisData <- read.csv('URL')
View(DennisData)
```

```
import pandas as pd
DennisData = pd.read_csv('URL')
DennisData.head(5)
```

```
options = weboptions('ContentType', 'table');
DennisData = webread('URL', options);
```

Click for Matlab NetCDF (slide 6)

What do the “I” and “R” in FAIR mean to you?

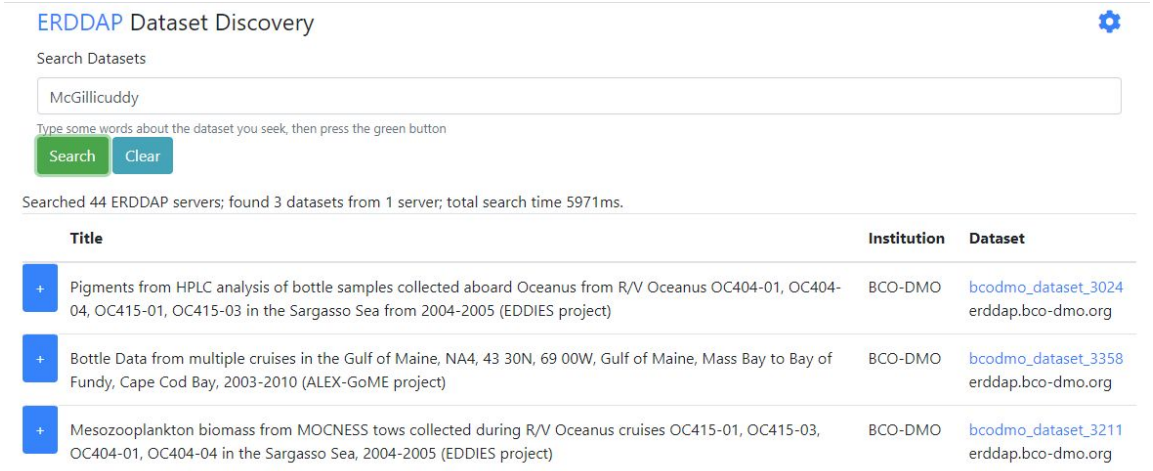
Did anyone pull data into an analysis tool?
Did you do this by manual download or API?
Did anyone inspect the metadata for interoperability?
Do you think you could reuse the data?

Check out data access options in other domain repositories, e.g.,:
PO.DAAC
EDI

Follow-up article for ERDDAP activity:
Signell, R.P.; Fernandes, F.; Wilcox, K. (2016) Dynamic Reusable Workflows for Ocean Science. *J. Mar. Sci. Eng.*, *4*, 68, https://doi.org/10.3390/jmse4040068

What about the “F” Findable and ERDDAP?

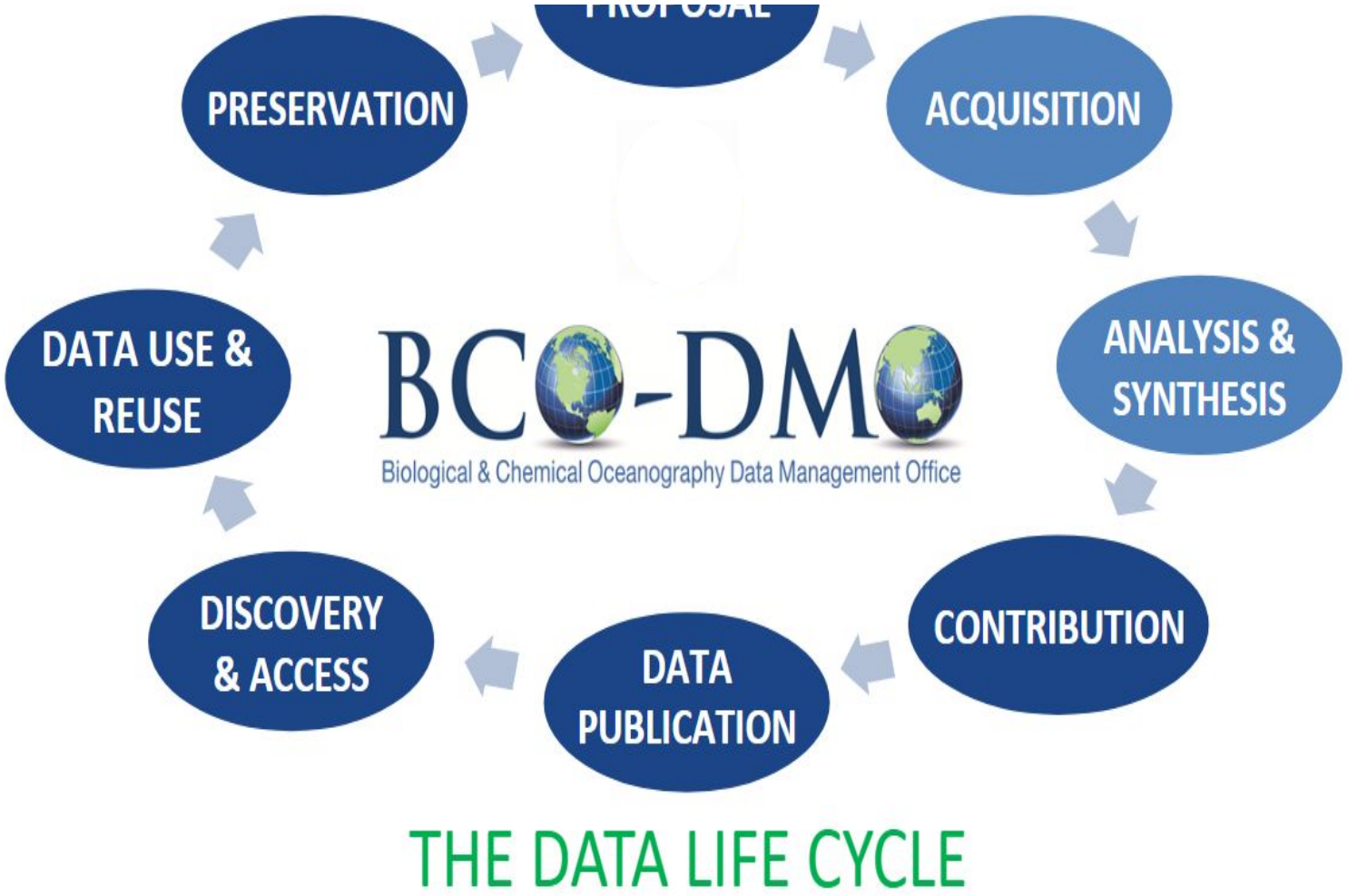
ERDDAP megasearch:
http://erddap.com/



https://coastwatch.pfeg.noaa.gov/erddap/download/SearchMultipleERDDAPs.html

https://github.com/IrishMarineInstitute/awesome-erddap

Congratulations! You completed WHOI's Data Science Training Camp! What's next?



4:15pm

Strategies to learn more

- Self-paced learning (online resources, books)
- Join and/or create communities
- Attend workshops in person

See Box 1 and Box 2 in:

Lowndes, J.S.S., et al. (2017) Our path to better science in less time using open data science tools. *Nature Ecology & Evolution,* 1, 0160, doi:10.1038/s41559-017-0160

Online resources: Data life cycle

- ESIP Data Management Training (DMT) Clearinghouse
  http://dmtclearinghouse.esipfed.org/
- DataONE Education Modules
  https://www.dataone.org/education
- Data Carpentry lessons
  https://datacarpentry.org/lessons/

Plus resources listed in Module 2 slides

Online resources: Software and computing

- Udemy
  https://www.udemy.com/
- Coursera
  https://www.coursera.org/browse/data-science
- Software Carpentry lessons
  https://software-carpentry.org/lessons/

Online resources tailored to our domain

- International Oceanographic Data and Information Exchange (IODE) OceanTeacher Global Academy https://classroom.oceanteacher.org/ (materials based on in-person workshops that you can apply to)
- OceanHackWeek: https://github.com/oceanhackweek/Oceans19-data-science-tutorial
- Integrated Marine Observing System (IMOS) (Australia) Marine Data and Science e-lectures https://open2u.utas.edu.au/Course/4261

Resources at MIT and in Boston area

- MIT Library is offering workshops, e.g., “Make your research computationally reproducible,” Introduction to Python on Jan. 28th (https://libraries.mit.edu/news/events/)
- MIT Statistics and Data Science Center https://stat.mit.edu/

  SDSCon 2020, MIT Statistics and Data Science Conference, Friday, April 3, 2020 @E14-674.
- Introduction to Computational Thinking and Data Science https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/

https://www.meetup.com/Boston-useR/

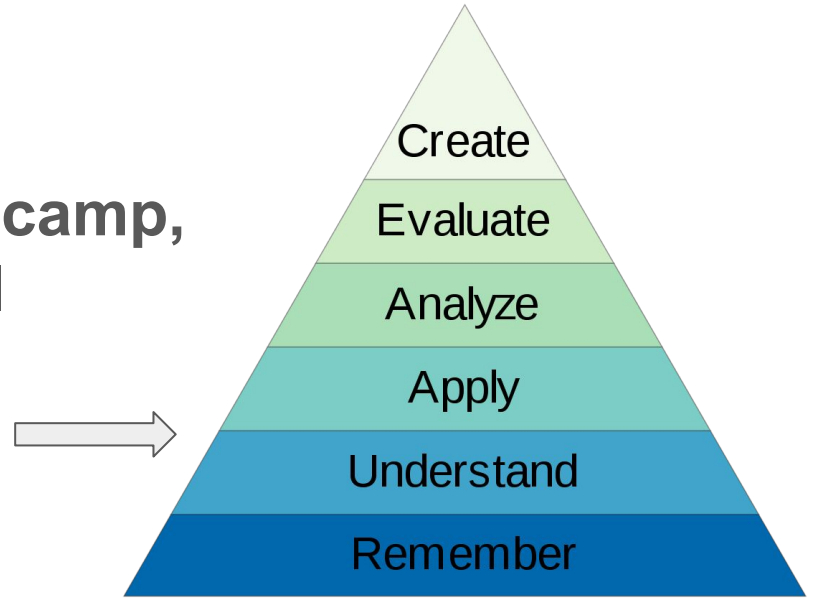
Resources in our Woods Hole community

- Woods Hole data-mongers email list
  http://mailman.whoi.edu/mailman/listinfo/data-mongers
- WHOI Software email list
  http://mailman.whoi.edu/mailman/listinfo/whoi-software
- Bioinformatics email list
  http://mailman.whoi.edu/mailman/listinfo/bioinformatics-users
- WHOI's Matlab users group (mug@whoi.edu)
- Woods Hole Python users group (whython@googlegroups.com)
- There is interest but no champion yet:
  - local R users group
  - local data science discussion group (*start with your lab group*)

We are planning a Summer Series with 4 “Lunch-and-Learns”...

“Exposure and Confidence” *

When you complete the camp, you may reach this level for some skills

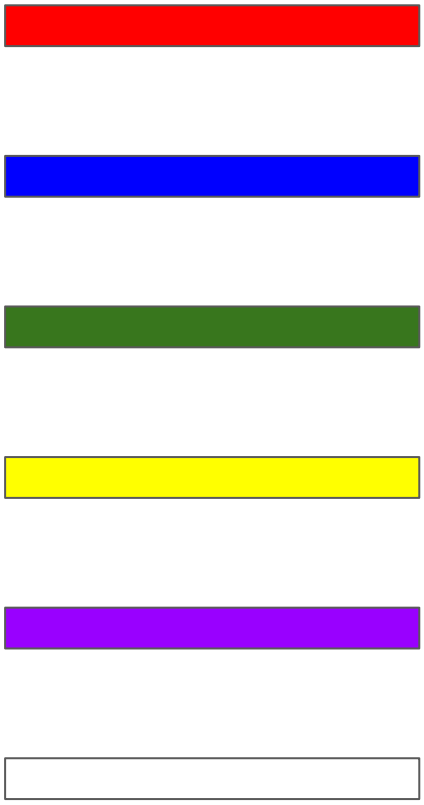


“Bloom's taxonomy” of levels of learning

For higher level learning, we will offer
The Carpentries Workshops...

* Lowndes, J.S.S., et al. (2017) Our path to better science in less time using open data science tools. *Nature Ecology & Evolution,* 1, 0160, doi:10.1038/s41559-017-0160

Please choose name tag by department.
Goal is for each table to have 1 or 2 depts.



AOPE

Biology

G&G

MC&G

PO

Other, or would rather not specify