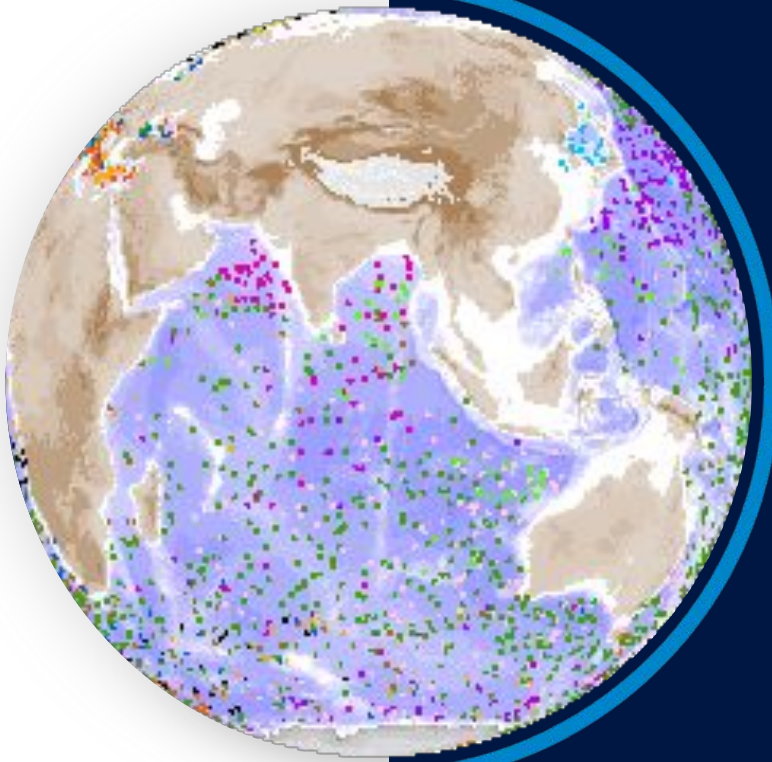
Data Accessibility

Katherine Silliman (presenter), Heather Holbach, Kathryn Sellwood, Brittany Troast, Luke Thompson, & Shaun Dolk

Atlantic Oceanographic & Meteorological Laboratory
National Oceanic and Atmospheric Administration
U.S. Department of Commerce

Table of Contents



Why do we care about being FAIR?

Findable – Assigned a globally unique and persistent identifier (DOI), indexed in a searchable resource.

Accessible – Openly retrievable and free to access.

Interoperable – Should use a formal, accessible, shared, and broadly applicable language for knowledge representation (**data standards**), machine-readable formats

Reusable – Richly described with accurate and relevant attributes, accessible data usage license, and meet domain-relevant community standards.

Why do we care about being FAIR?

Federal requirement for intra- and extra-mural NOAA funded research

Findable

Accessible

Interoperable

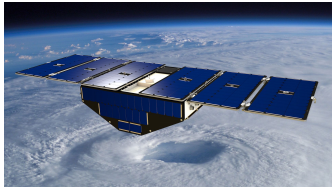
Reusable

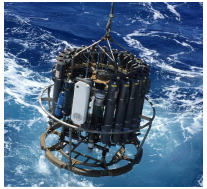
NOAA Plan for Increasing Public Access to Research Results

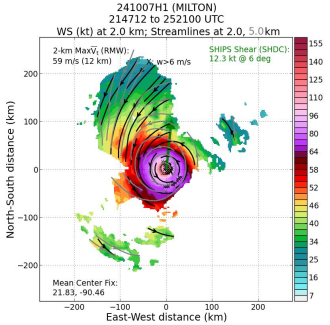
A Response to the White House Office of Science and Technology Policy Memorandum *Increasing Access to the Results of Federally Funded Scientific Research* issued February 22, 2013

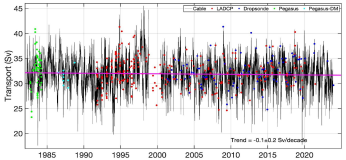
AOML data are as varied as our research portfolio

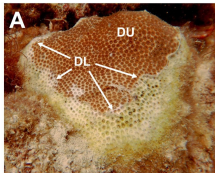
Observations

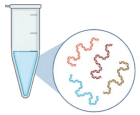




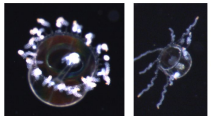




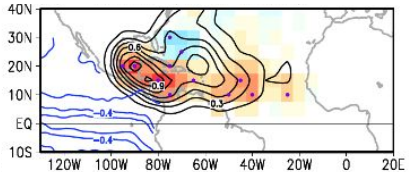


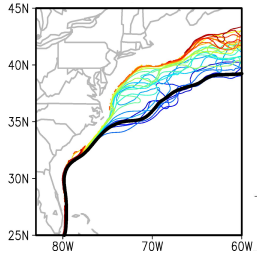


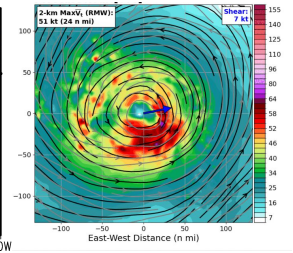
Gelatinuos



Models







Software

Tourmaline

Amplicon Sequence Processing Workflow

github.com/aomlomics/tourmaline

NOAA-AOML

Reports

Probabilities of 2024 NATL seasonal hurricane activity

Relative to 1951-2020 June-November climatology

Below-normal	Near-normal	Above-normal
5%	5%	90%

AOML Data Management Resources

Added Since Last Review

Computing and data storage

Leveraging on-premise and cloud computing and storage, including new data center

On-premise data management support

Provides long-term public access to AOML data and products and facilitates the submission of AOML data to NOAA archives

Engagement with NOAA and external data practice groups

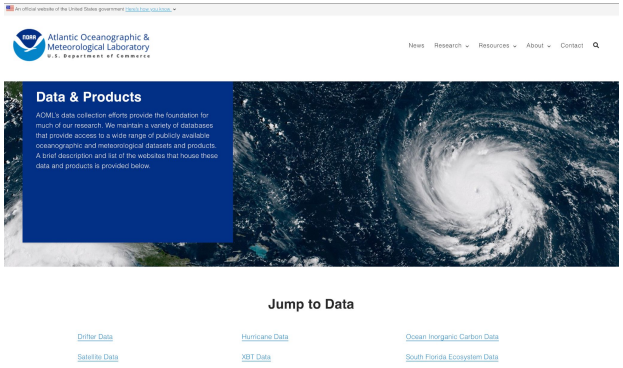
Participates in working groups internally and externally to discuss and develop data best practices

Transition Specialist

Facilitates collaboration with end users and stakeholders to ensure data requirements are met

AOML Data Accessibility

AOML website



Data-specific Repositories

NCBI

OBIS OCEAN BIODIVERSITY INFORMATION SYSTEM

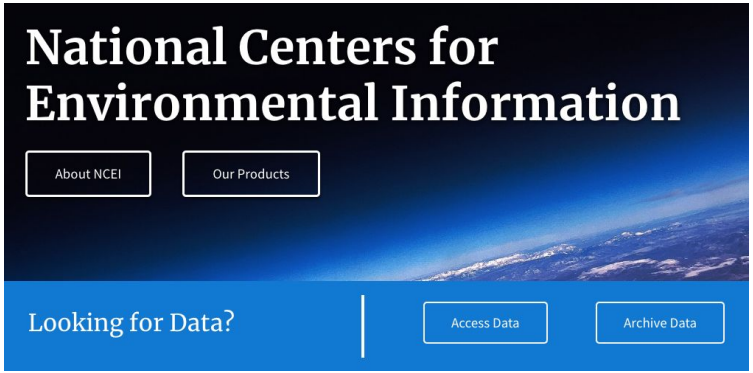
AcDC

GLOBAL DRIFTER PROGRAM

SOCAT SURFACE OCEAN CO2 ATLAS

Argo

NOAA's NCEI Archive and Data.gov



Data Assembly Centers

Cloud Storage and Archive

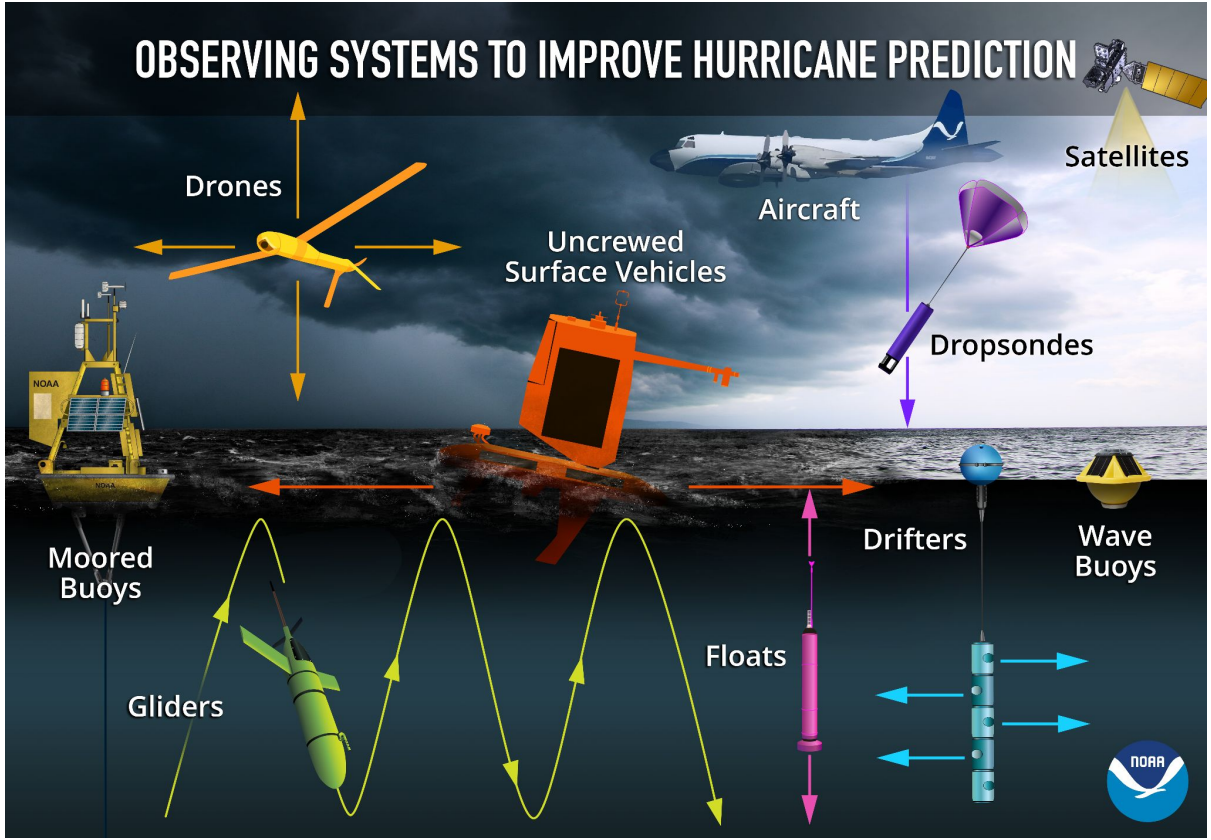
Direct Reports to Stakeholders

Primary Data Users

- Other NOAA labs, line offices, and programs (e.g., Weather Service, Fisheries, Sanctuaries, Coastal Science)
- Academic institutions
- Local, regional, and national resource management stakeholders
- Military applications (including Search and Rescue)
- Private industry (e.g., insurance, construction, agriculture, maritime safety and navigation, oil and gas)

Hurricane Data

- Air and Sea Platforms
- Near real-time global transmission
- Multi-agency collaborations



Hurricane Data

Hurricane Helene (AL09)

Monday, September 23 – Thursday, September 26

Flight Info and Data

Wind Centers | Track | Track Plot

– 20240923U1

Monday, September 23

Take Off: 1426Z Biloxi	**Aircraft**: AF309	**Tasking**: NHC
Landing: 2347Z Biloxi	**Mission ID**: 0109A	**Pattern**: Invest

Download Data:

SFMR	Dropsonde	Flight-Level (ASCII)	

Mission Documents:

ARWO Report

– 20240925I1

Tuesday, September 25

Take Off: 0906Z Lakeland	**Aircraft**: NOAA43	**Tasking**: NHC
Landing: 1612Z Lakeland	**Mission ID**: 1009A	**Pattern**: Butterfly

Download Data:

SFMR	Dropsonde	Flight-Level (NetCDF)	Flight-Level (text)

Mission Documents:

FltDir Report | FltDir Log | LPS log

+ 20240924U1

+ 20240924N1

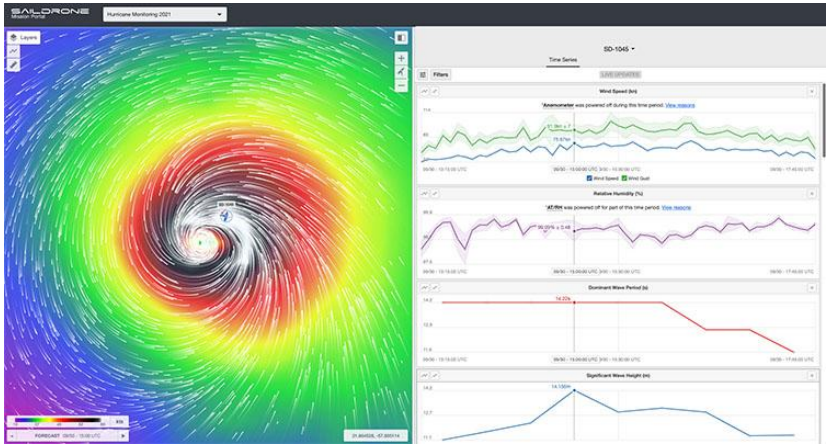
+ 20240925U2

+ 20240925U4

https://www.aoml.noaa.gov/data-products/#hurricanedata

- Access to research quality standard aircraft data is made available as soon as possible.
- Only public location to obtain full resolution data from the U.S. Air Force Reserve Hurricane Hunters!

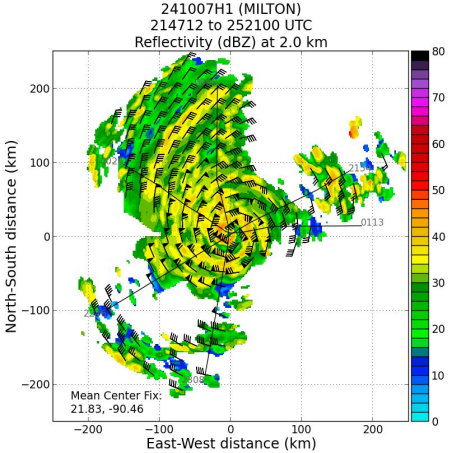
Saildrone Dashboard - Hurricane Sam (2021)

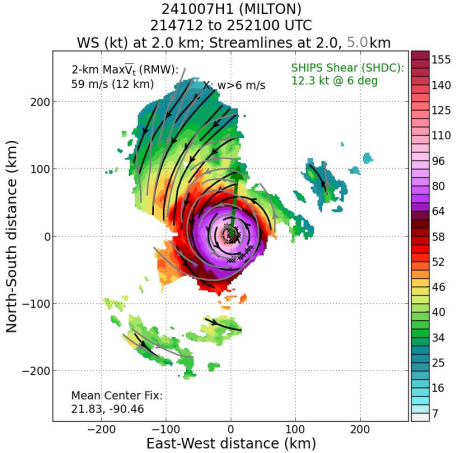


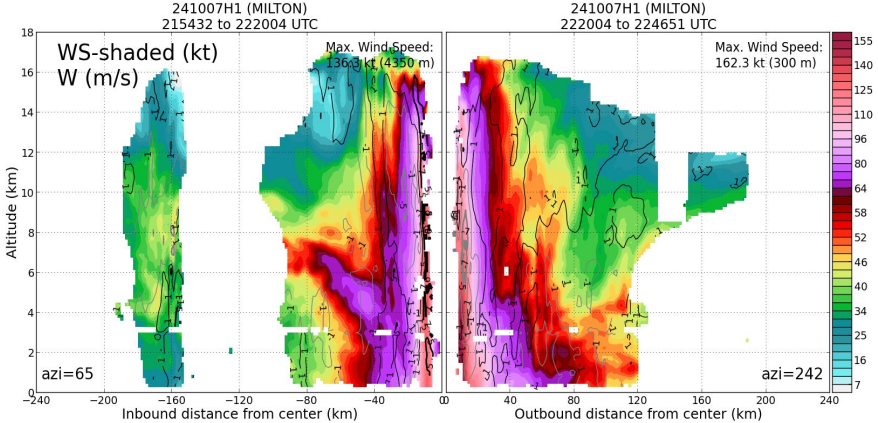
A Selection of Available Graphics and Data

Real-Time Graphics

Tail Doppler Radar







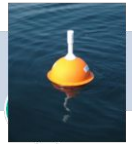
These graphics and raw data can be found at:

https://www.aoml.noaa.gov/ftp/hrd/data

Ocean Sampling Time Series

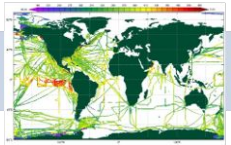
EXpendable BathyThermograph (XBT) Network 1960

1979 Global Drifter Program (GDP)



Western Boundary Time Series (WBTS) 1982

1991 Ocean Carbon Cycle (OCC)

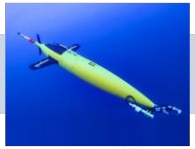


US Argo Data Assembly Center (DAC) 1999

2006 PIRATA Northeast Extension (PNE)

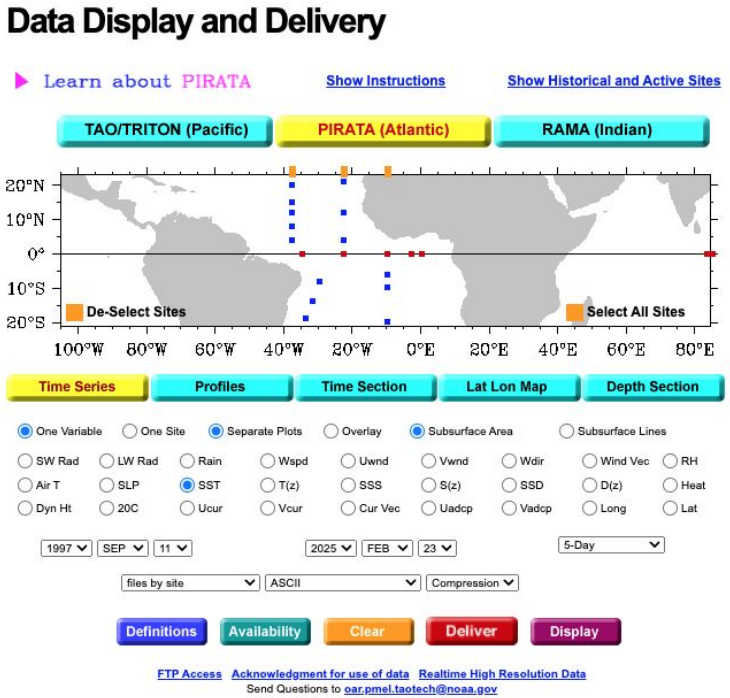
Southwest Atlantic Meridional Overturning Circulation (SAM) 2009

2014 Underwater Glider

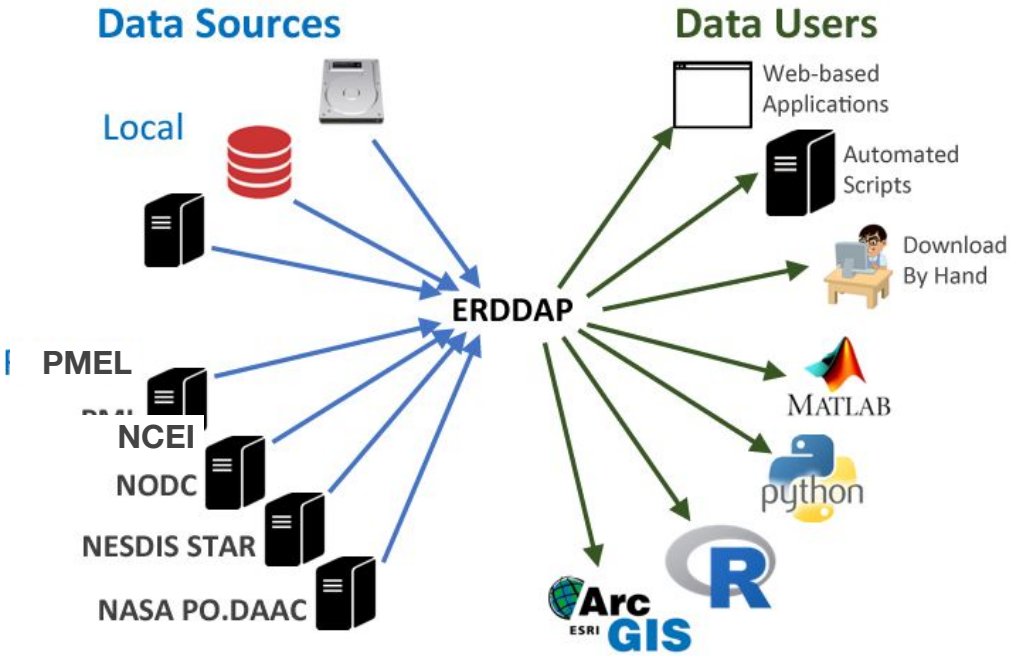


Hurricane Saildrone 2021

Ocean Sampling Data



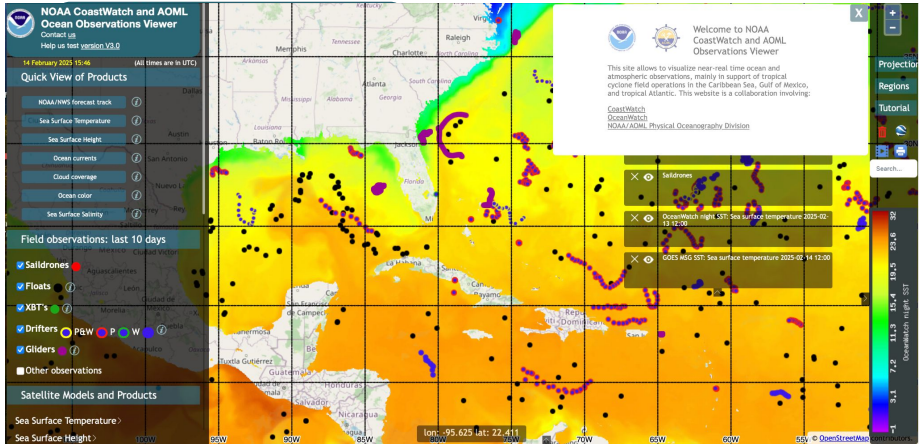
Dashboards provide interactive, ‘code-free’ access historical and ongoing data products



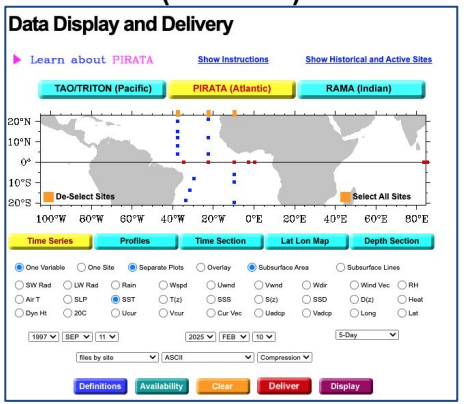
ERDDAP data servers support automated data retrieval for data analysts, web developers, and modelers

Autonomous Data in ‘Real time’

CoastWatch Ocean Viewer

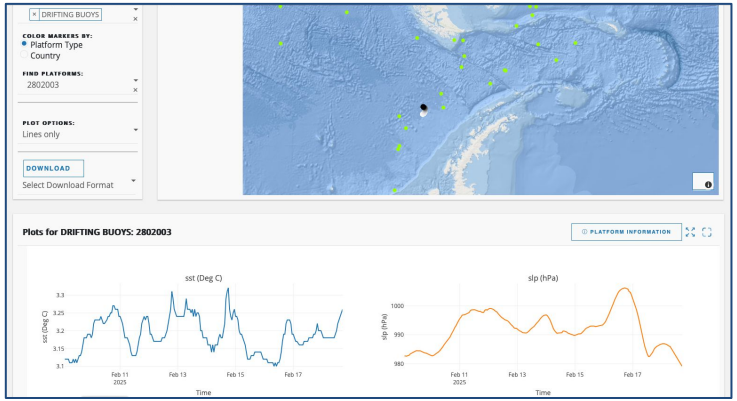


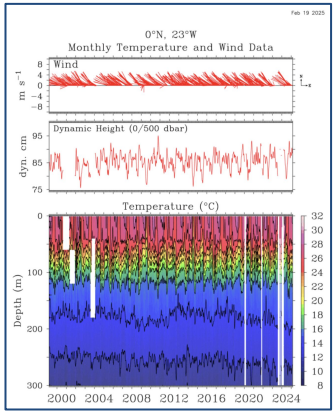
Tropical Moored Buoy Array (PIRATA)



Array Map

NOAA Observing System Monitoring Center (OSMC)

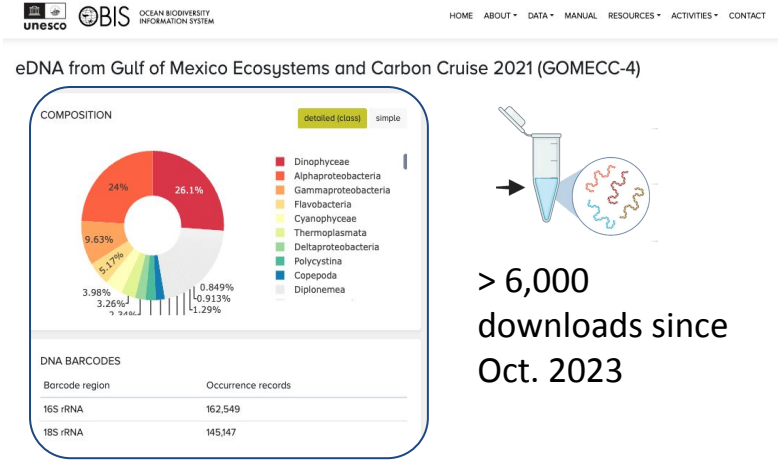




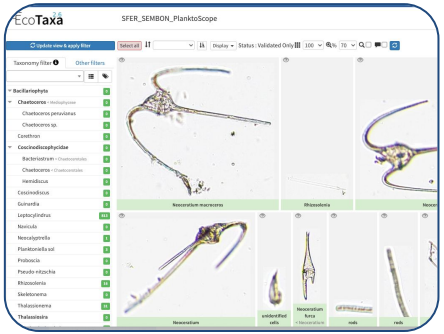
Station Summary Report

Ecosystem Data

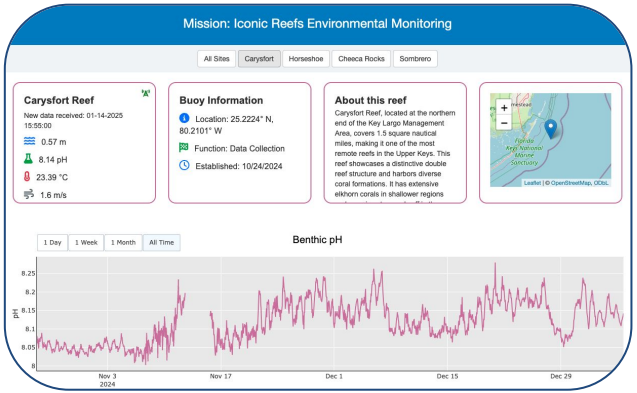
Biodiversity data shared to global platforms



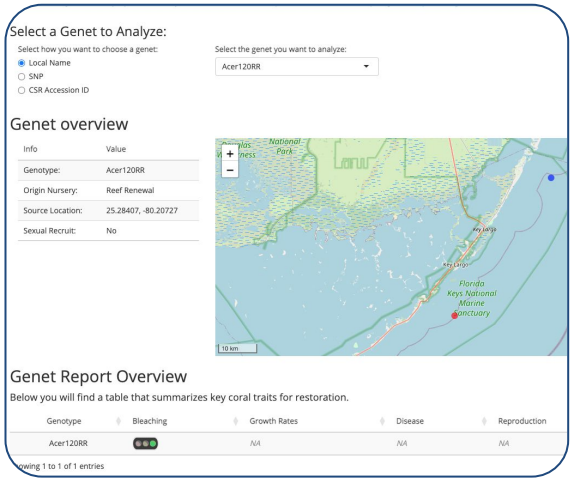
> 6,000 downloads since Oct. 2023



PlanktoScope images shared to EcoTaxa



Real-time data dashboard for Florida Keys Sanctuary buoys



Decision support tool for coral restoration (ACDC)

Ecosystem Assessments

Data

Long-term collaborative monitoring projects

- Comprehensive Everglades Restoration Plan

Integrated Ecosystem Assessments (IEA)

- Comprehensive Everglades Restoration Plan

Open-Science Framework

Openscapes

- Professional Development

GitHub

- Host data and data visualization products

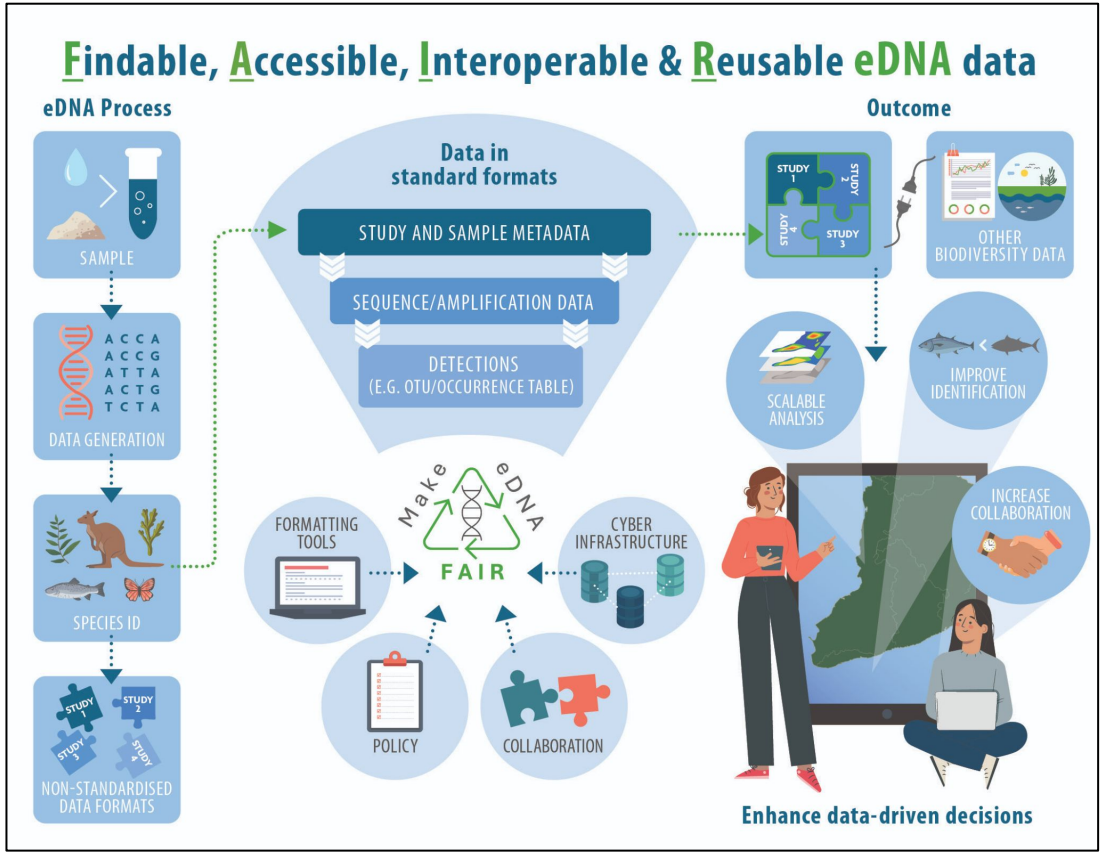
Accessibility & Automation

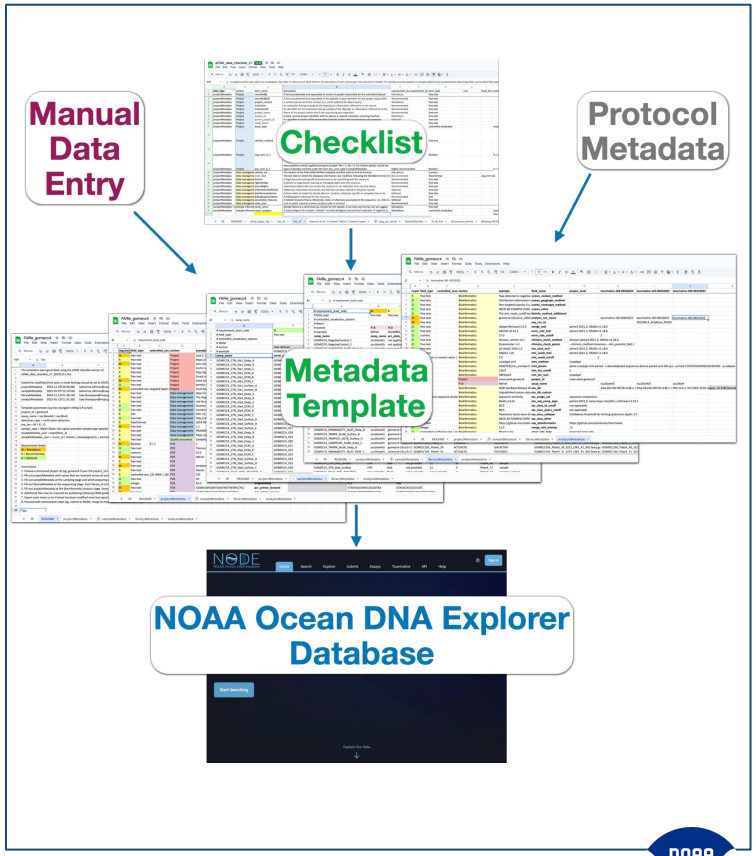
Hosting Data

- NCEI
- ERDDAP

Automated reports and websites

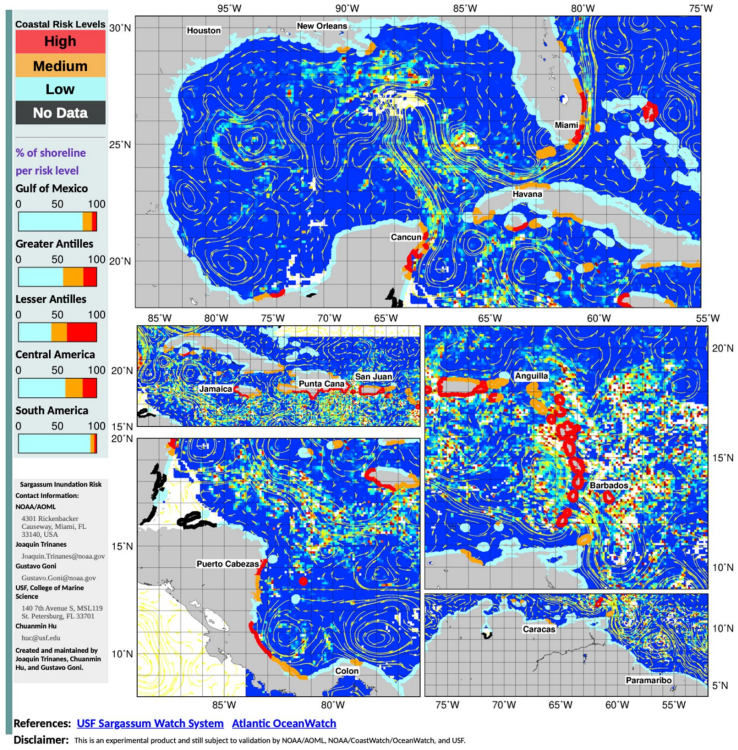
Making ‘Omics Data FAIR



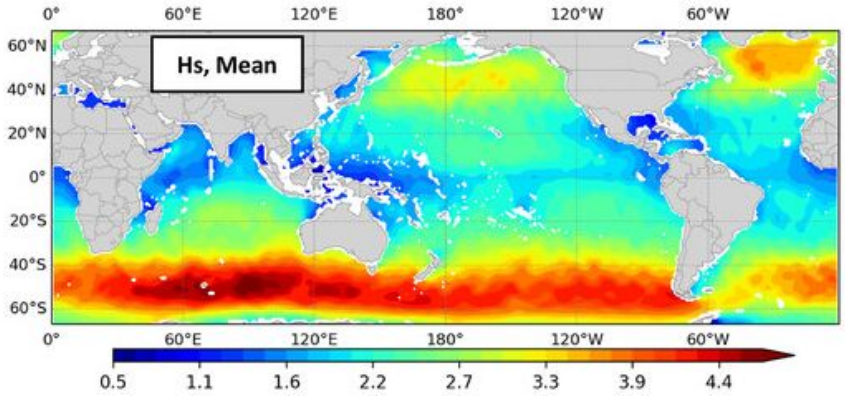


https://fair-edna.github.io

Model outputs



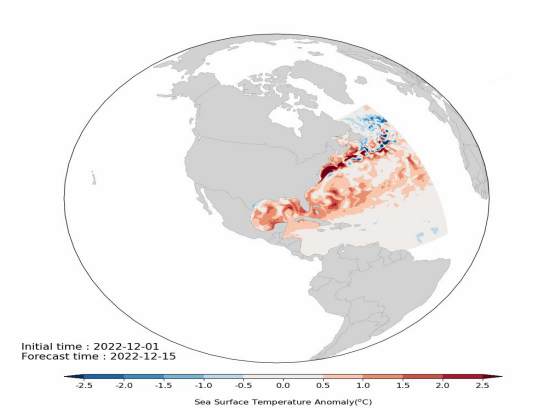
Experimental Weekly Sargassum Inundation Risk



2° x 2° Mean Significant Wave Height

Fig. 17(c)
Campos, R.M, et. al., (2024)

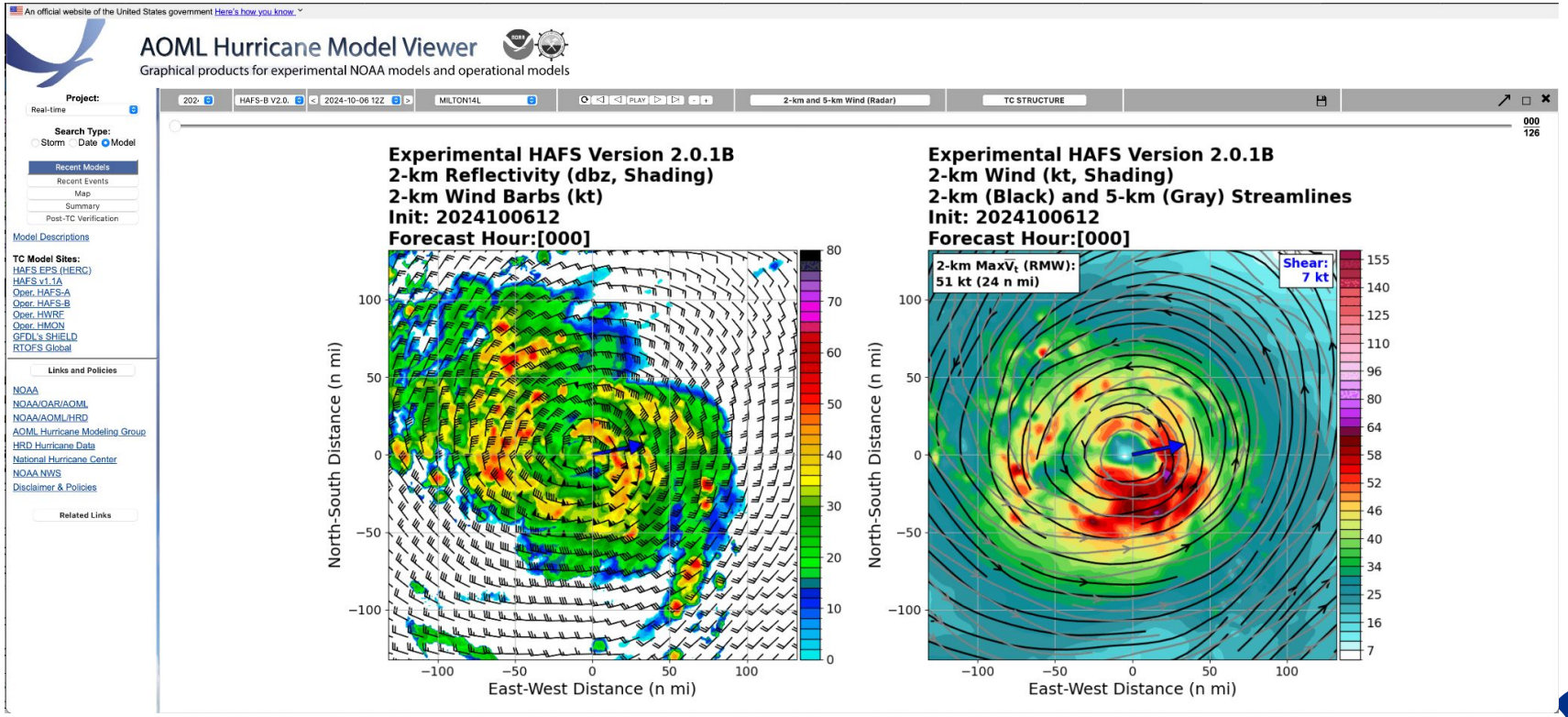
Global Wave Ensemble Forecast



CEFI portal

(led by NOAA/PSL)

Hurricane Model Viewer



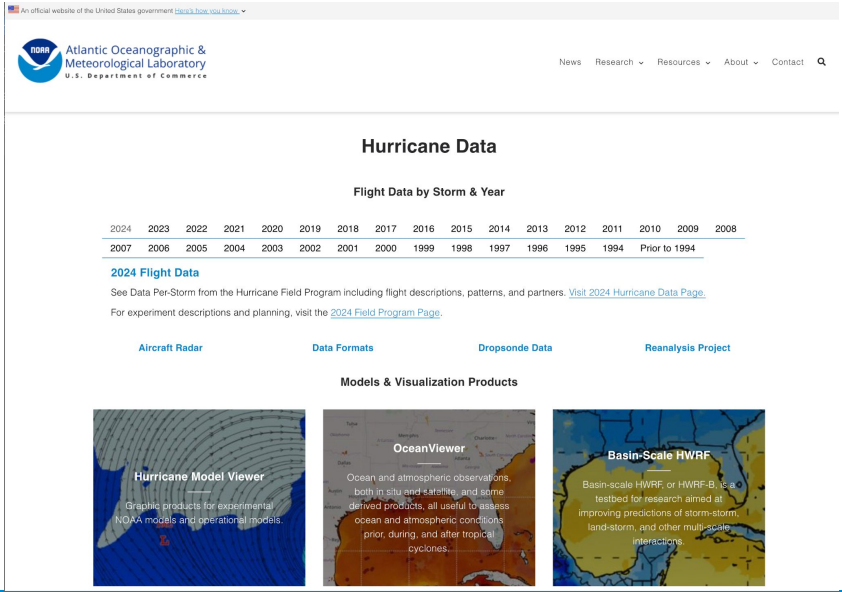
Looking ahead

- FINDABLE: Obtaining DOIs for all datasets
- ACCESSIBLE:
  - Testing new ways to improve data user experience
  - Improving speed of data publication
- INTEROPERABLE:
  - Digitizing historical data archives (from 1960!)
  - Developing and adopting community data standards
- REUSABLE: Providing comprehensive metadata on how data was collected and processed

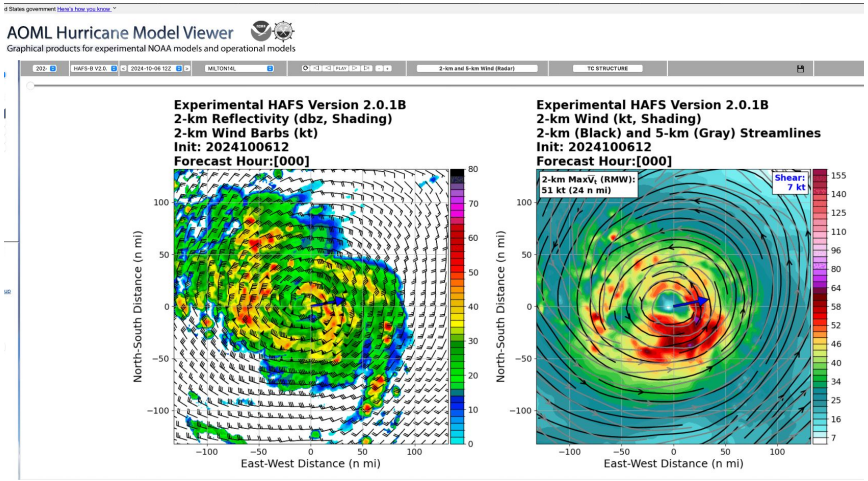
FAIR data practices take lots of time and expertise, but ultimately increases the return on investment for our research!

Supplemental slides for reference

Hurricane Research Division



Hurricane Data



Hurricane Model Viewer

Data is available via ftp server and linked to the aoml web page with https access. https://www.aoml.noaa.gov/data-products/#hurricanedata

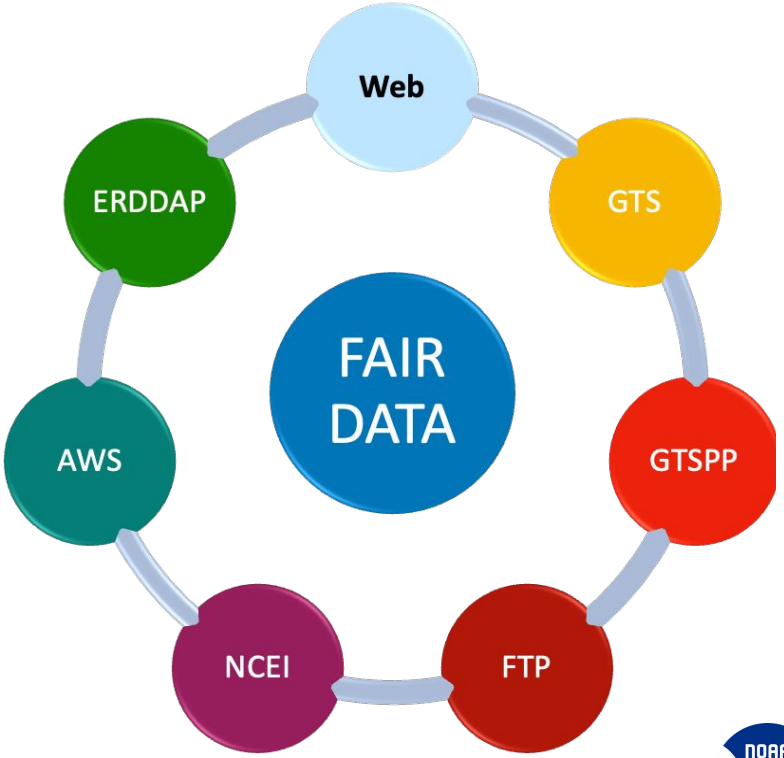
Operational and research model products

Ocean Sampling - Recently Developed Hindcast Datasets

- **Surface ocean-biogeochemical patterns derived from a 0.10-degree resolution MOM5-TOPAZ hindcast:** 1958 - 2019 (FA. Gomez, SK. Lee (2023))
- **Surface North Atlantic MOM5-TOPAZ outputs derived from a regular hindcast and a robust diagnostic simulation experiment:** 1980 - 2017 (FA. Gomez, SK. Lee (2023))
- **GoMBio-model derived surface and bottom fields of temperature, salinity, total alkalinity, and dissolved inorganic carbon in the Gulf of Mexico:** 1950 - 2020 (FA. Gomez, SK. Lee (2023))
- **RC4USCoast - A river chemistry dataset for regional ocean model application in the U.S. East, Gulf of Mexico and West Coasts:** 1950 - 2020 (FA. Gomez, et al., (2022)
- **Surface patterns of temperature, salinity, total alkalinity (TA), and dissolved inorganic carbon (DIC) across the Gulf of Mexico derived from the GoMBio model experiments:** 1981 - 2014 (FA. Gomez, R. Wanninkhof, L. Barbero, (2022))

Ocean Sampling - Data Accessibility (PhOD)

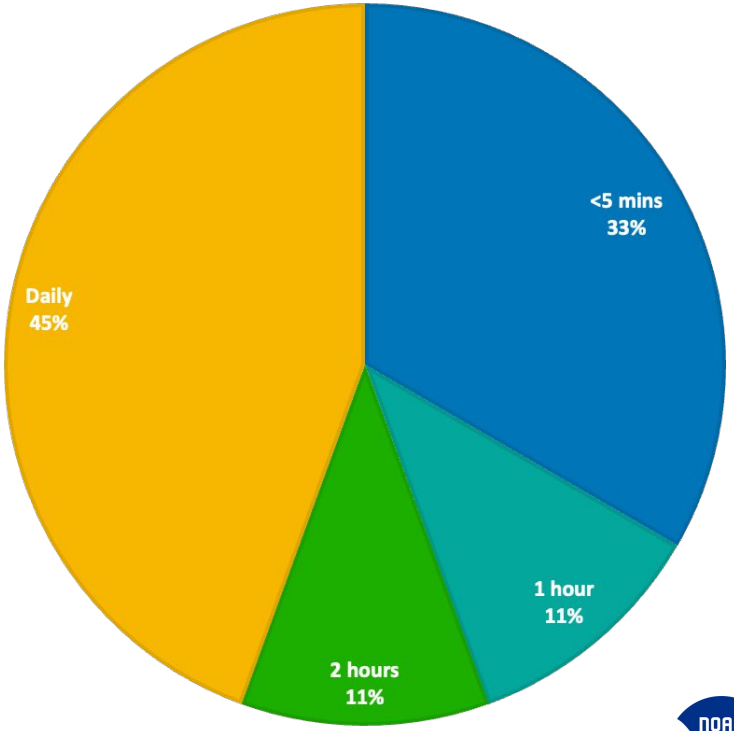
- Data available from many sources, in a variety of file formats. Including but not limited to:
  - Web
    - Data.gov
    - Project Web Pages
    - Data Acquisition Centers (DACs) and Global Data Acquisition Centers (GDACs)
    - NOAA CoastWatch
  - Global Telecommunication System (GTS)
  - Global Temperature and Salinity Profile Programme (GTSPP)
  - FTP
  - NOAA's NCEI (3 DOIs, 2 pending)
  - Amazon Web Services (AWS)
  - ERDDAP
    - Local and Federated Systems



Ocean Sampling - Data Timeliness (PhOD)

- 33% of projects provide data <5 mins
- 50% of projects provide data within 2 hours
- 100% of projects provide data within 1 day of transmission!

Raw Data Access by Project



- Quality Control Data Availability
  - Nearly all projects provide quarterly QC updates
  - QC updates are cruise dependant for 3 projects

Ecosystem Assessments

Goal

All projects, big or small, produced in an **open-science framework**, hosted in an **online data portal** that facilitates customized report creation, data visualization dashboards, and a**ccessibility to diverse stakeholders**

In Development

IEA R Package

ESR Dashboards

Example