Bridging NOAA Ocean Model Files and Cloud-Native Access

Using Virtual Stores and Zarr-backed services to simplify access and unlock scientific workflows

Eli Holmes, PhD • NOAA Fisheries
Lead, NOAA Fisheries Open Science • ESIP July 2026

Gridlock + Icechunk (Wed), Openscapes (Wed), Federal Open Data policy, OM hackweeks

NOAA ocean model products are core national infrastructure for navigation, public safety, management, and planning.

NOAA Operational Forecast Systems
Daily guidance for safe navigation, port operations, and coastal decision-making.
CBOFS, CIOFS, DBOFS, GoMOFS, NYOFS, TBOFS, WCOFS, SSCOFS, SFBOFS, NGOFS2, LEOFS, LMHOFS...

Regional Coastal Forecast Systems
Large-area forecasts used to prepare for changing ocean and coastal conditions.
ECCOFS, NECOFS, SECOFS

CEFI Regional Ocean Models
Regional MOM6 models used to support fisheries and marine ecosystems management

Coastal Ocean Reanalysis
Long-term reconstructions used to understand past conditions and extreme events.

National Water Model Retrospective
Decades of modeled river and stream conditions used for flood analysis and research.

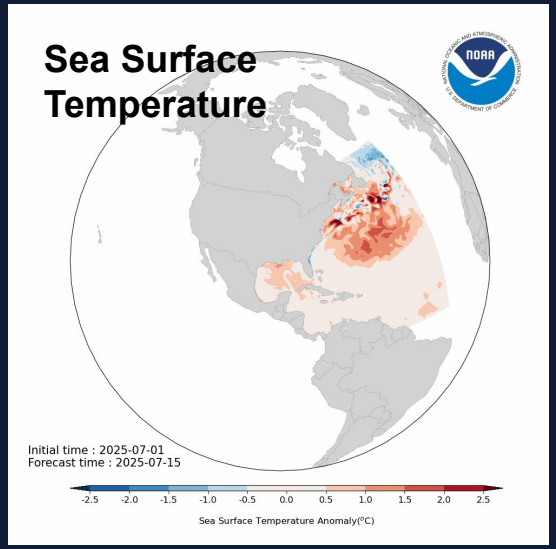
Data Volume

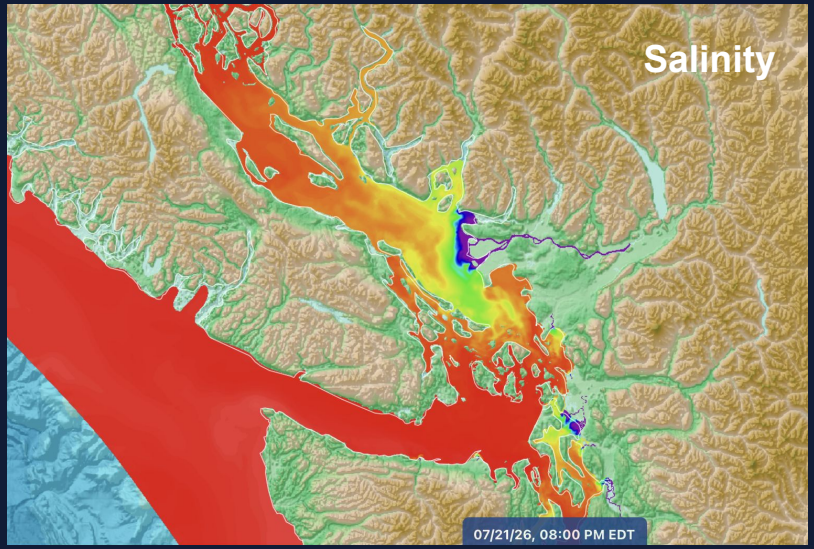
Massive Output

The models behind NOAA's ocean, coastal and water services produce an **enormous volume** of data.

CORE NOAA OPERATIONAL DATA PRODUCTS

Ocean Model Data: Examples





🔗 Link to CEFI Data Portal

🔗 Link to OceanMap

🔗 NOAA Tide and Currents

So many netcdf files!



Example from CEFI: MOM6 regional models for fisheries and ocean ecosystems

chlos.nep.full.hcast.daily.regrid.r20250912.202501-202512.nc	7 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199301-199312.json	8 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199301-199312.nc	8 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199401-199412.json	8 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199401-199412.nc	8 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199501-199512.json	8 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199501-199512.nc	8 months ago
dissic.nep.full.hcast.daily.regrid.r20250912.199601-199612.json	8 months ago

Showing 1 to 50 of 820 entries

This massive volume of data is generated from just a single configuration:

Just One Regional Model

One Hindcast

One Region & One Grid

One Release

Challenges & Bottlenecks

Common User Problems

Server Bottlenecks

THREDDS/OPeNDAP servers are problematic when data are enormous and user requests are large.

Users are resorting to downloading raw data locally

The File Barrier

Navigating so many files is a big barrier for users:

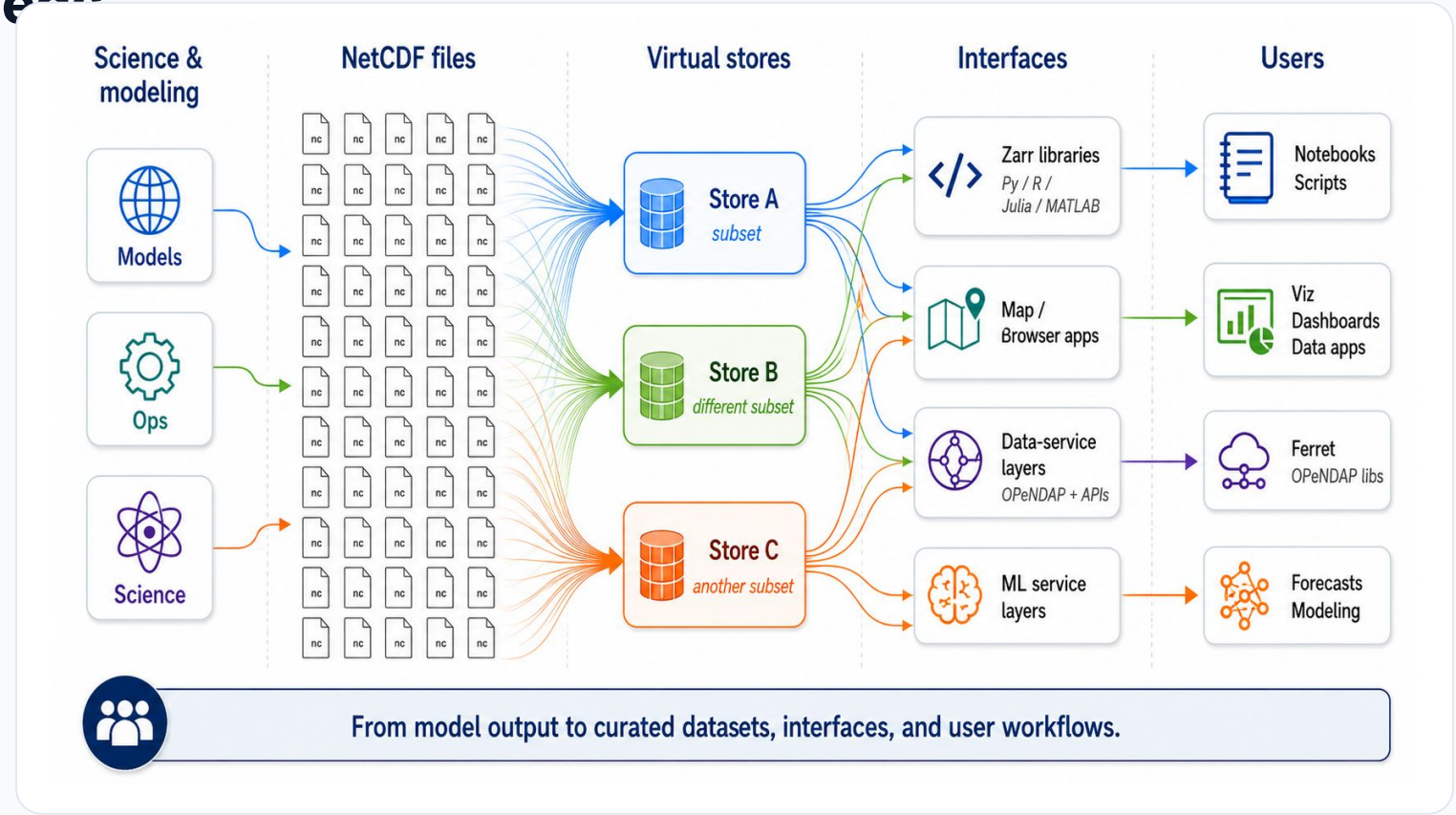
Complex Discovery What files to use? What variables?

Time Series Overhead Creating time series requires many files.

Visualization Friction Data visualization requires many files and writing code.

Data Virtualization

Virtualization provides a solution for meeting user needs



Two NOAA implementations of these virtualization pipelines for ocean model data

NOAA/NOS/IOOS

Cloud-Optimized Data Management & Cyberinfrastructure (CO-DMAC)

An established end-to-end implementation for operational forecast models (OFS), coastal operational forecast models (COFS), coastal ocean reanalysis (CORA), and other operations models.

Today's Focus

NOAA/OAR/PMEL

Climate, Ecosystems, and Fisheries Initiative (CEFI)

Regional MOM6 models supporting fisheries and marine ecosystem management. Implemented using ArrayLake from Earthmover—the example I will focus on today.

IOOS CO-DMAC contacts and information

Technical Questions

Kelly Knee
kelly.knee@tetratech.com

Jonathan Joyce
jonathan.joyce@tetratech.com

Project POC

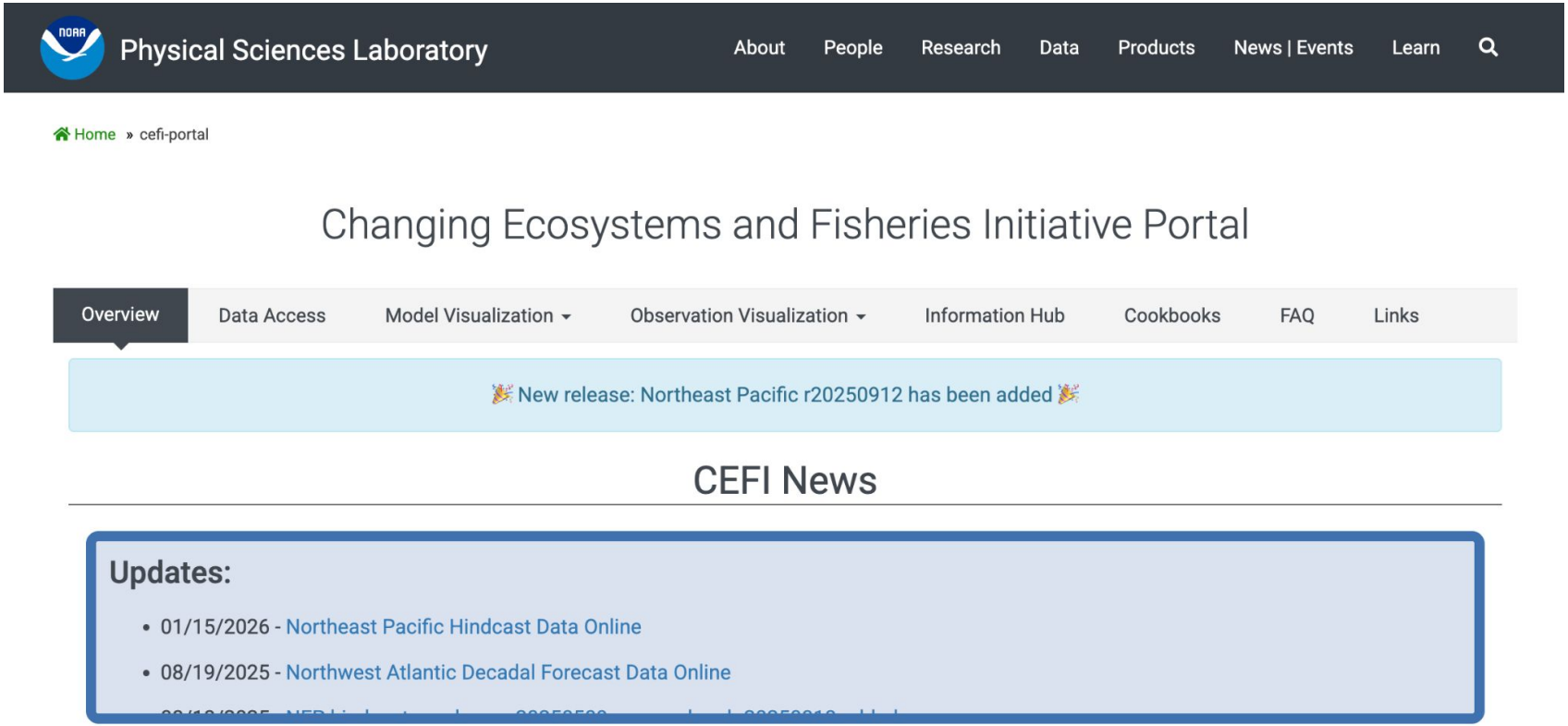
Micah Wengren
micah.wengren@noaa.gov



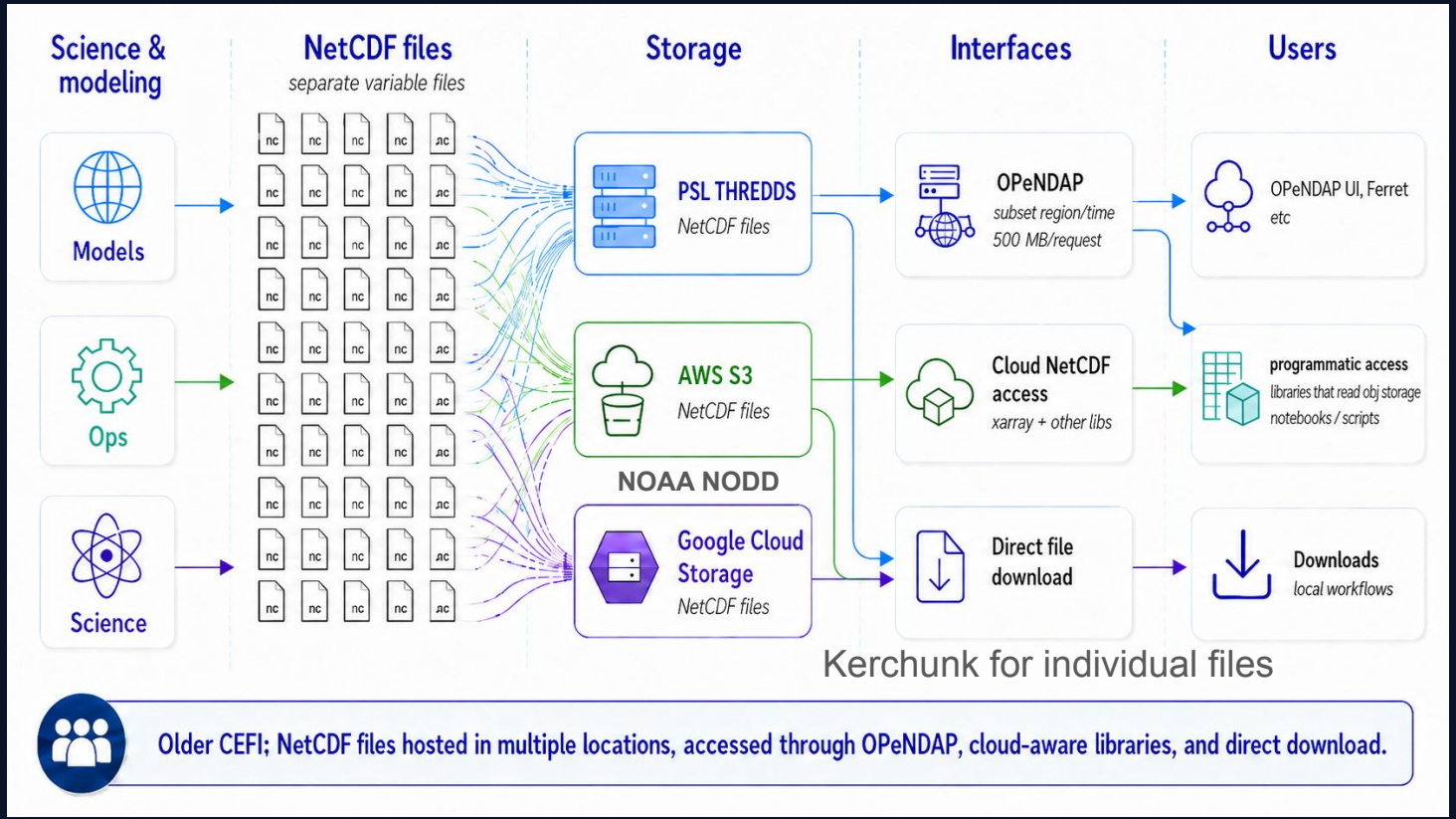
Annual Meetings & Resources

IOOS DMAC Annual Meetings
2026 IOOS DMAC Annual Meeting Presentation
2025 IOOS DMAC Annual Meeting Presentation

https://psl.noaa.gov/cefi_portal/



Old pipeline



Challenges & Bottlenecks

Familiar User Problems

Users are currently NOAA / NOAA partners but products will be operational later

Server Bottlenecks

THREDDS/OPeNDAP servers are problematic when data is enormous and user requests are large.

Users are hitting the 500MB limit.

⚠ **Users are resorting to downloading raw data locally**

The File Barrier

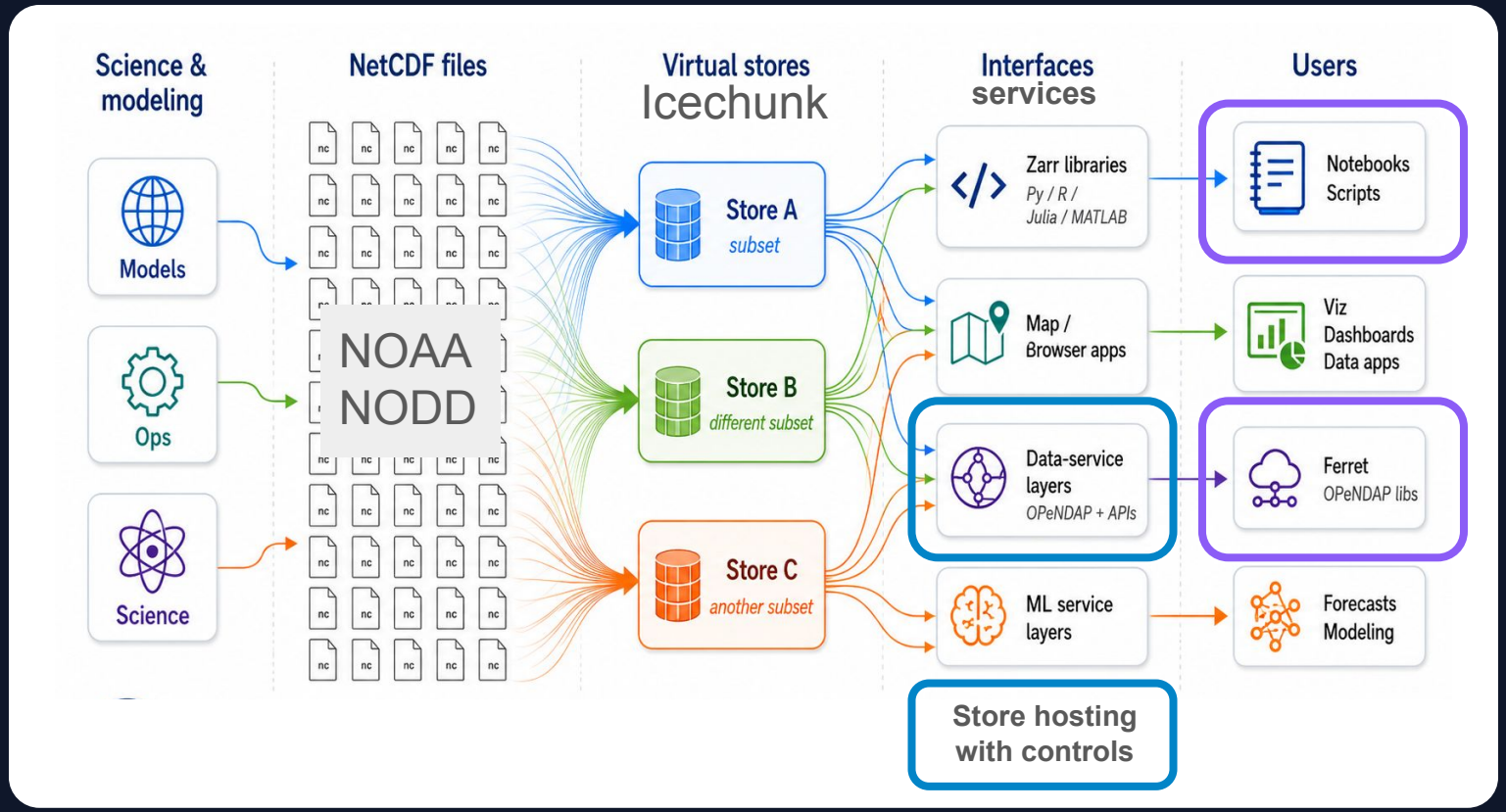
Navigating so many files is a big barrier for users:

Complex Discovery
What files to use? What variables?

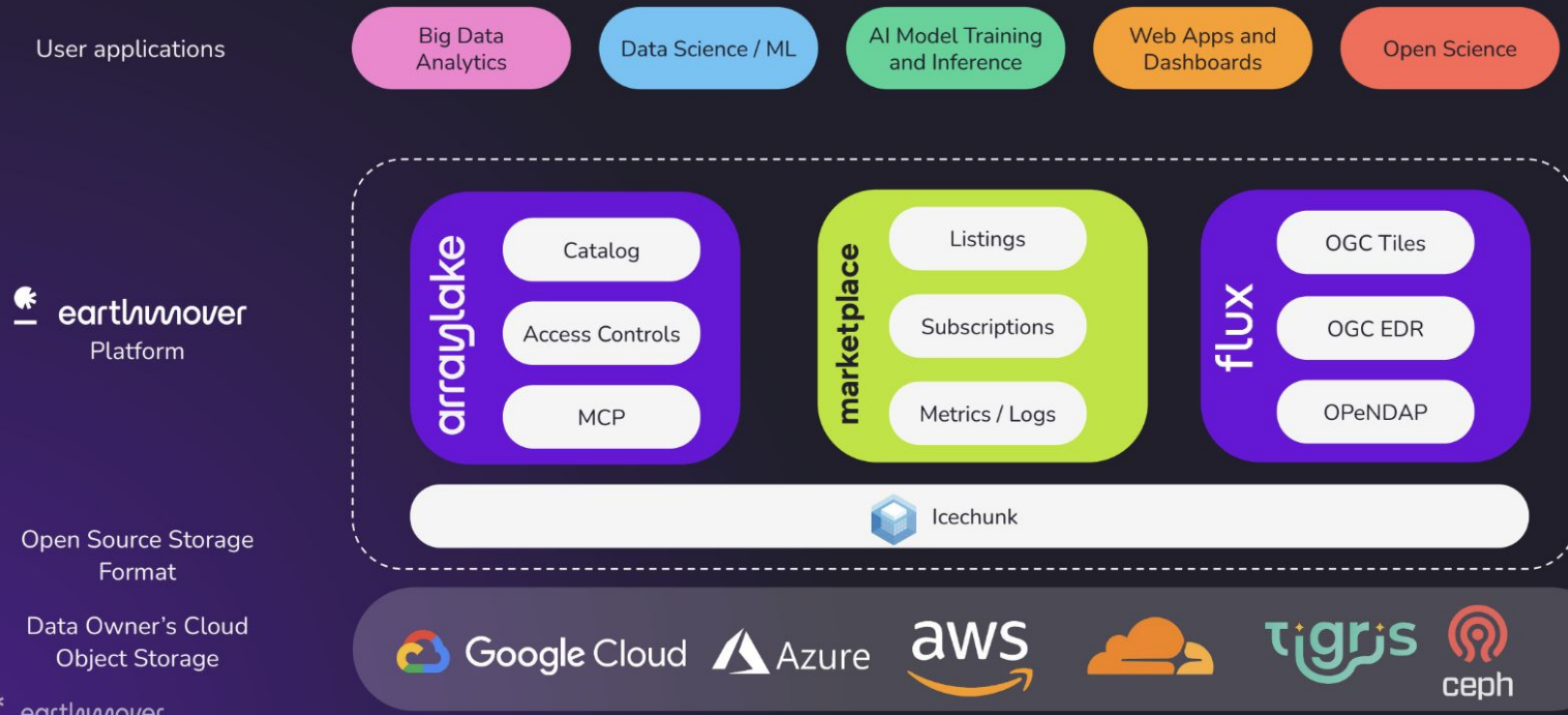
Time Series Overhead
Creating time series requires many files.

ARCO Data Set Creation is Hard
Creating Analysis Ready Datasets requires long, time-consuming scripts (not easily reproducible or sharable).

New pipeline



Earthmover: The ARCO Data Platform



Slide from Joe Hamman, Earthmover

"Users were asked to give examples of where they were running into problems."

Performance Comparison

Super Simple Request using OPeNDAP endpoint:

Full domain, one variable, all times, 1 depth (850+MB)

PSL THREDDS url hits server-side limits

```
library(ncdf4)
url =
"http://psl.noaa.gov/thredds/dodsC/Projects/CEFI/regional_mom6/cefi_portal/northeast_pacific/full_domain/hindcast/monthly/regrid/r20250912/vo_rotate.nep.full.hcast.monthly.regrid.r20250912.199301-202506.nc"
nc = nc_open(url)

vo_surface <- ncvar_get(
  nc,
  varid = "vo_rotate",
  start = c(1, 1, 1, 1),
  count = c(-1, -1, 1, -1)
)
```

FAILS

Earthmover ArrayLake doesn't

```
library(ncdf4)
url <-
"https://compute.earthmover.io/v1/services/dap2/NOAA-PMEL/cefi-nep-hindcast-monthly/main/regrid/main/opendap"
nc <- nc_open(url)

vo_surface <- ncvar_get(
  nc,
  varid = "vo_rotate",
  start = c(1, 1, 1, 1),
  count = c(-1, -1, 1, -1)
)
```

3 minutes

Performance Comparison

Eliminating downloading and data munging

From user example notebook

Download data

Script assumes you have these files in a `OcModels/Data/MOM6` folder. To download necessary inputs, run:

Depth (2-d)

- *deptho* `wget -O ocean_static.deptho.nc https://psl.noaa.gov/thredds/fileServer/Projects/CEFI/regional_mom6/cefi_derivati`

Mixed layer depth (2-d)

- *MLD_003* `wget -O MLD_003.nep.full.hcast.monthly.regrid.r20250509.199301-202412.nc https://psl.noaa.gov/thredds/fileServer/Projects/CEFI/regional_mom6/cefi_portal/n 202412.nc`

Temperature (3-d)

- *thetao* `wget -O thetao.nep.full.hcast.monthly.regrid.r20250509.199301-202412.nc https://psl.noaa.gov/thredds/fileServer/Projects/CEFI/regional_mom6/cefi_portal/n 202412.nc`

X & Y sea water velocities (3-d)

- *uo_rotate* `wget -O uo_rotate.nep.full.hcast.monthly.regrid.r20250509.199301-202412.n https://psl.noaa.gov/thredds/fileServer/Projects/CEFI/regional_mom6/cefi_portal/n 202412.nc`
- *vo_rotate* `wget -O vo_rotate.nep.full.hcast.monthly.regrid.r20250509.199301-202412.n`

Load & Analyze (Python)

Load variables

```
ds_mld = xr.open_dataset("../input/MLD_003.nc")
ds_temp = xr.open_dataset("../input/thetao.nc")
# Loaded instantly from cloud storage
```

Analysis

```
Tpart = annual_mean_3d(ds_temp, ds_depth, "thetao", "Tpa
```

```
---Tcop---
5549406720 records with dimensions (384, 52, 815, 341)
5549406720 records with dimensions (384, 52, 815, 341)
160013568 records with dimensions (96, 52, 94, 341)
160013568 records with dimensions (96, 52, 94, 341)
Runtime: 12.44 minutes
---Tpart---
5549406720 records with dimensions (384, 52, 815, 341)
5549406720 records with dimensions (384, 52, 815, 341)
```

Performance Comparison

Same code but with data from the virtual store

No downloading •

```
client = Client()
repo =
client.get_repo("NOAA-PMEL/cefi-nep-hindcast-mo
nthly")
session = repo.readonly_session(branch="main")
aux = xr.open_zarr(session.store,
zarr_format=3, group="regrid/aux1")

# The double brackets extract these as dataset
# to keep the same structure as previous code

ds_mld = aux[['MLD_003']]

main = xr.open_zarr(session.store,
zarr_format=3, group="regrid/main")

ds_temp = main[['thetao']]
ds_xv = main[['uo_rotate']]
ds_yv = main[['vo_rotate']]
```

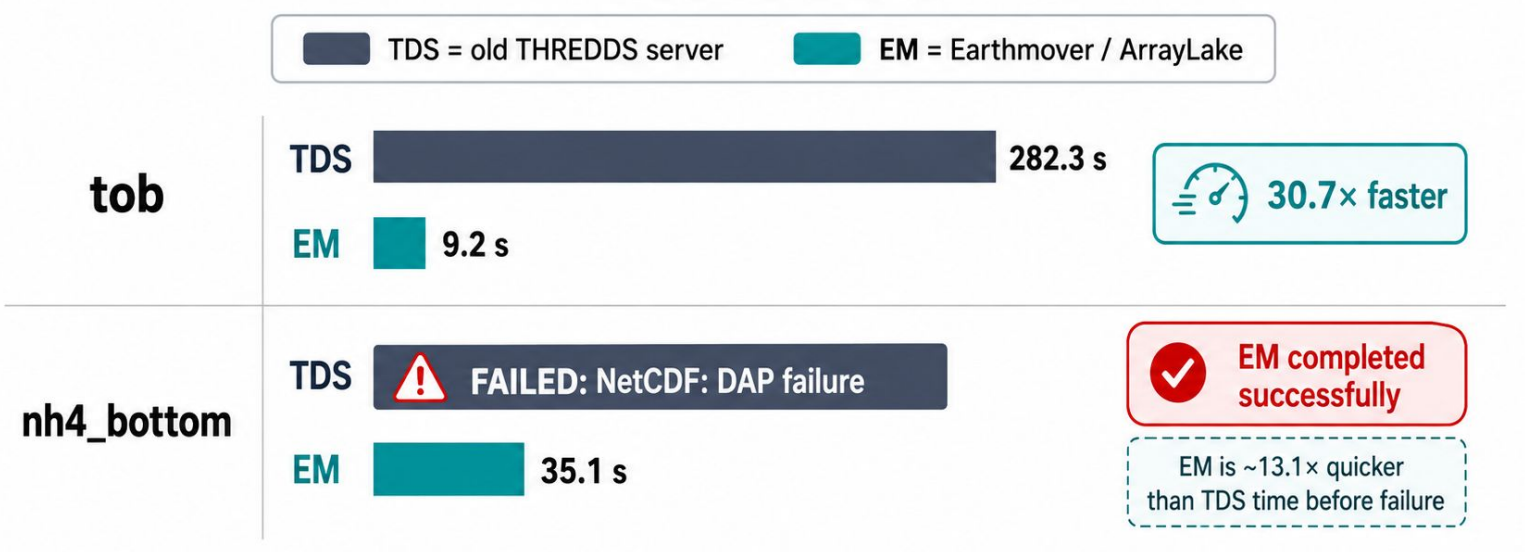
No code changes required

```
# temperatures
Tcop = annual_mean_3d(ds_temp, ds_depth, "thetao", "T
Tpart = annual_mean_3d(ds_temp, ds_depth, "thetao", "
```

```
---Tcop---
5636116200 records with dimensions (390, 52, 815, 341)
5636116200 records with dimensions (390, 52, 815, 341)
160013568 records with dimensions (96, 52, 94, 341)
160013568 records with dimensions (96, 52, 94, 341)
Runtime: 5.61 minutes
---Tpart---
5636116200 records with dimensions (390, 52, 815, 341)
5636116200 records with dimensions (390, 52, 815, 341)
220018656 records with dimensions (132, 52, 94, 341)
220018656 records with dimensions (132, 52, 94, 341)
Runtime: 5.11 minutes
```

Resampling a 2D MOM6-NEP variable

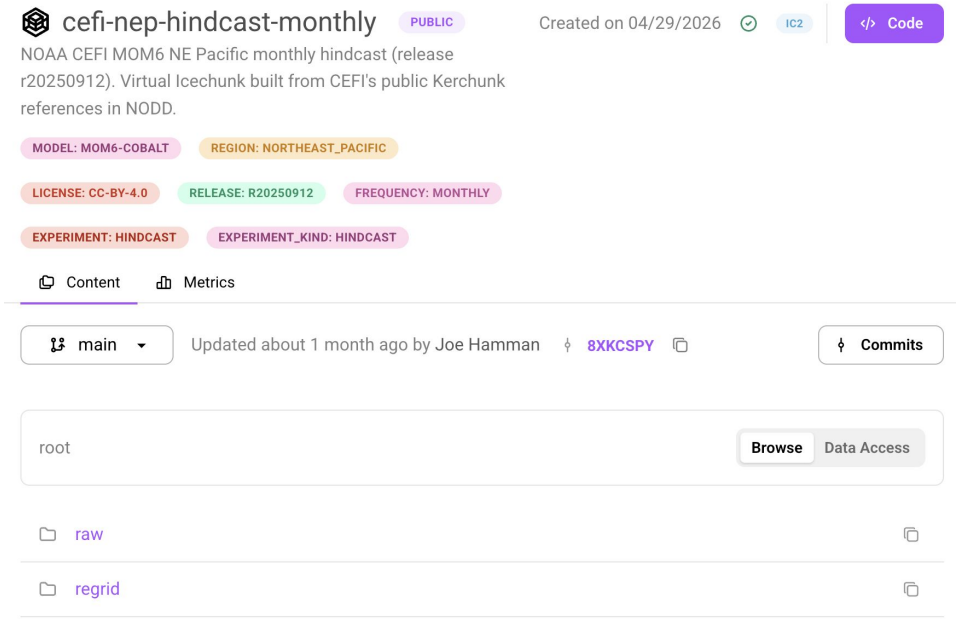
Performance Speed Ups: From minutes to seconds

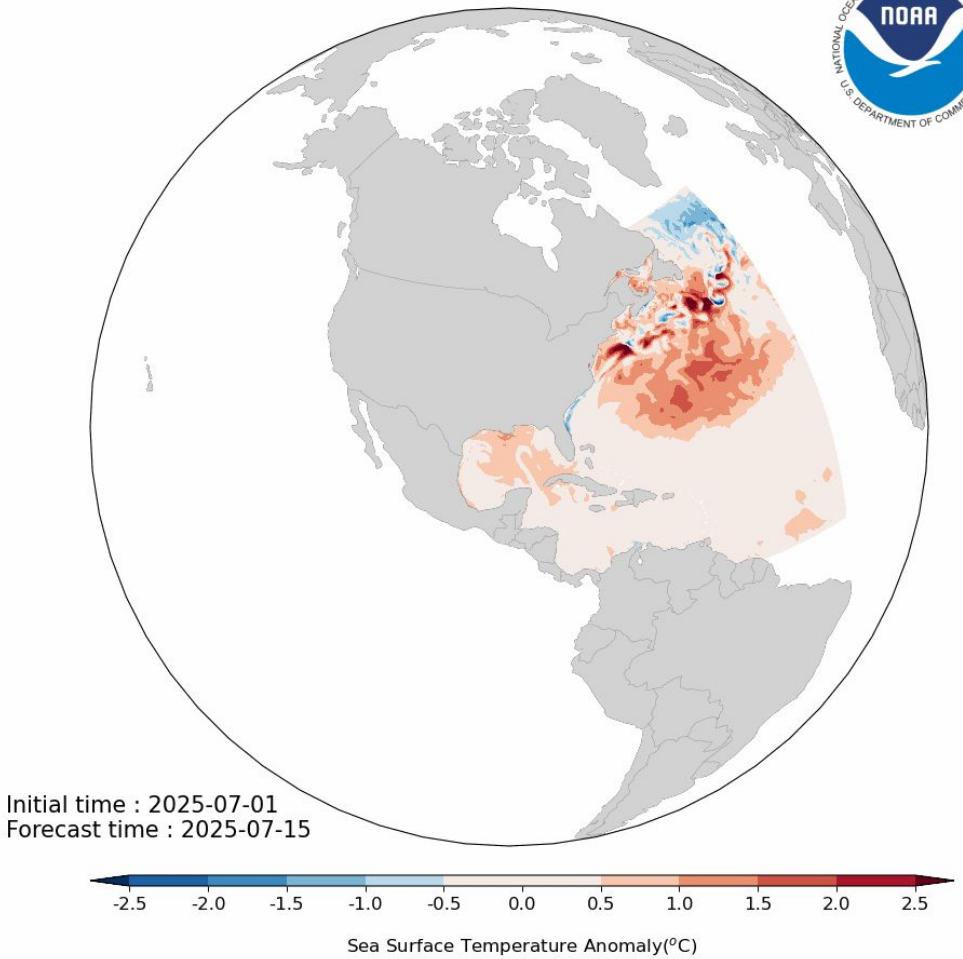


Comparable to the older ROMS experience, despite MOM6's substantially larger grid. Important since data consumers are used to the older fast ROMS experience.

Experiment with CEFI model output in ArrayLake

- https://app.earthmover.io/NOAA-PMEL/cefi-nep-hindcast-monthly
- https://app.earthmover.io/NOAA-PMEL/cefi-nep-hindcast-daily
- https://app.earthmover.io/NOAA-PMEL/cefi-nwa-decadal-forecast-monthly
- https://app.earthmover.io/NOAA-PMEL/cefi-nwa-decadal-forecast-yearly
- https://app.earthmover.io/NOAA-PMEL/cefi-nwa-hindcast-daily
- And more…





Thank you!

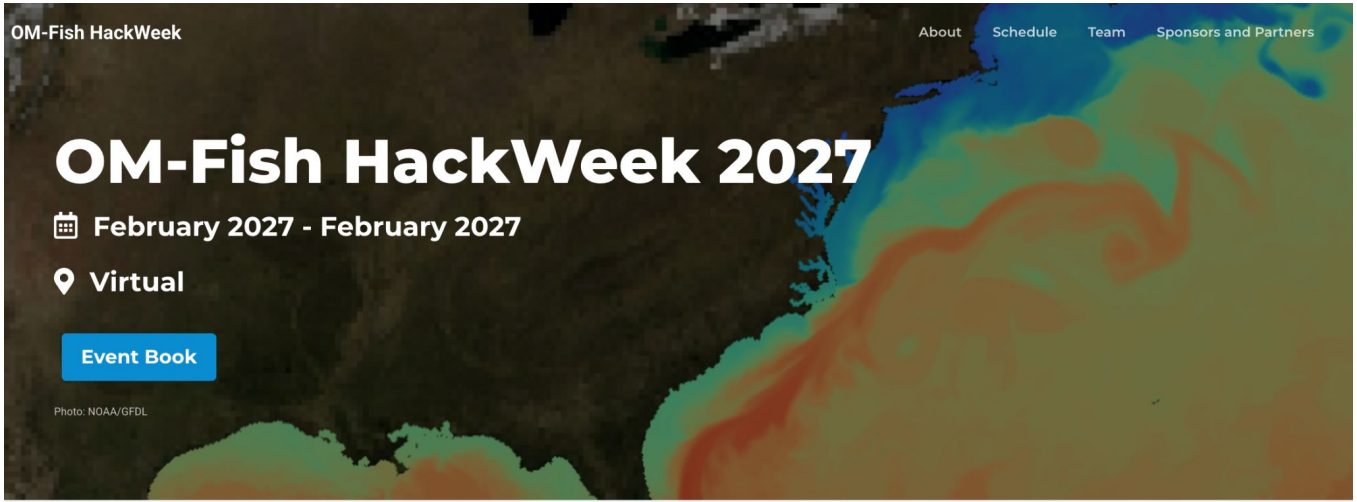
eli.holmes@noaa.gov

Questions re CEFI? Look for *Roland Schweitzer at ESIP*

Questions re ArrayLake and Earthmover? *Look for Joe Hamman and the Earthmover booth*

Questions re IOOS CO-DMAC and other NOAA ocean model data virtual stores? *Contact micah.wengren@noaa.gov*

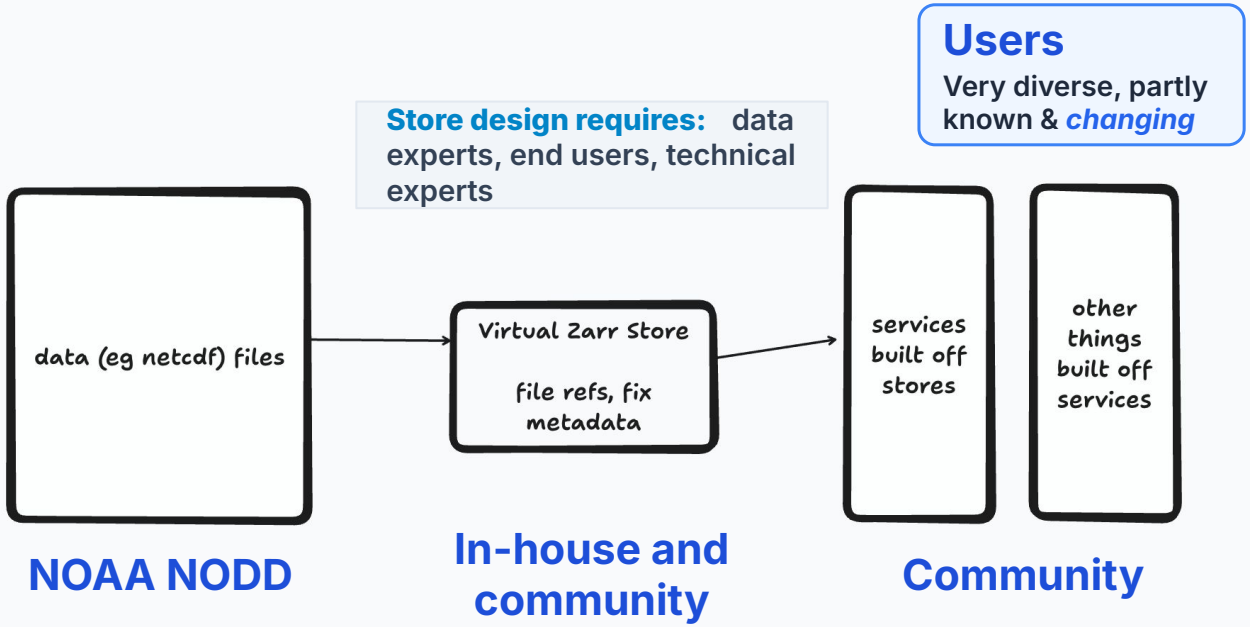
Do you love hackweeks and ocean model data? Reach out!



About OM-Fish HackWeek

Ocean models and reanalysis products—including regional systems such as ROMS and MOM6, global ocean reanalyses such as GLORYS, regional products such as MOM6-COBALT-NEP, and climate projections from CMIP6—are transforming our ability to study changing ocean and fisheries habitats across the globe. These products provide rich three-dimensional information about ocean temperature, oxygen, currents, mixed layer depth, and ecosystem conditions that can be used to investigate marine heatwaves, habitat compression, climate-driven range shifts, and changing ecosystem dynamics. In the OM-Fish HackWeek, you will learn to access, visualize, and analyze

No single virtual store meets all needs. Facilitate community developed virtual stores.



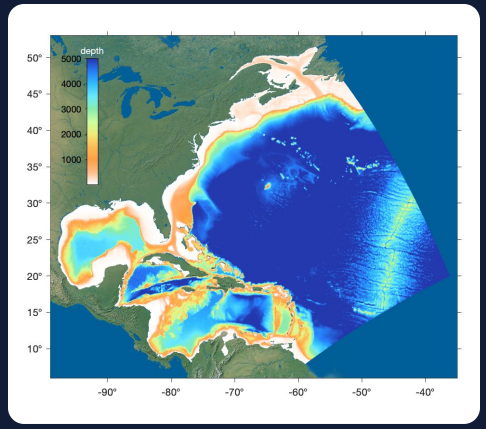
Ease of access → Increase in data reads from object storage (“digital wake”)

Operational Forecast Systems

Ocean Model Data: Examples

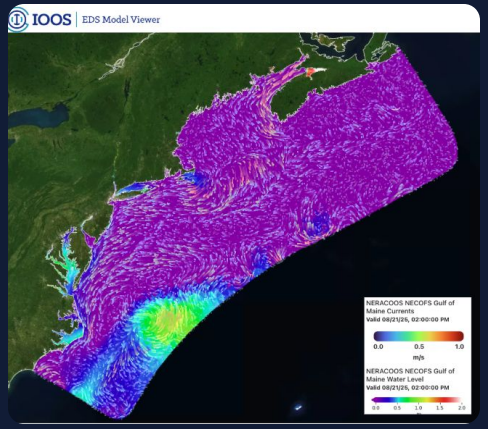
ECCOFS

East Coast Community Ocean Forecast System



NECOFS

NorthEast Coastal Ocean Forecast System



GoMOFS

Gulf of Maine Operational Forecast System

