HFRNet in the NCCF - An Overview of Services to Support Cloud-based Applications

Inger Kittle, Melissa Zweng, Mark Middlebusher

National Environmental Satellite, Data, and Information Service

04/29/2025

Outline

Project Overview

NCCF Framework

NCCF Operational Capabilities Today

Common Services and HFRNet Status

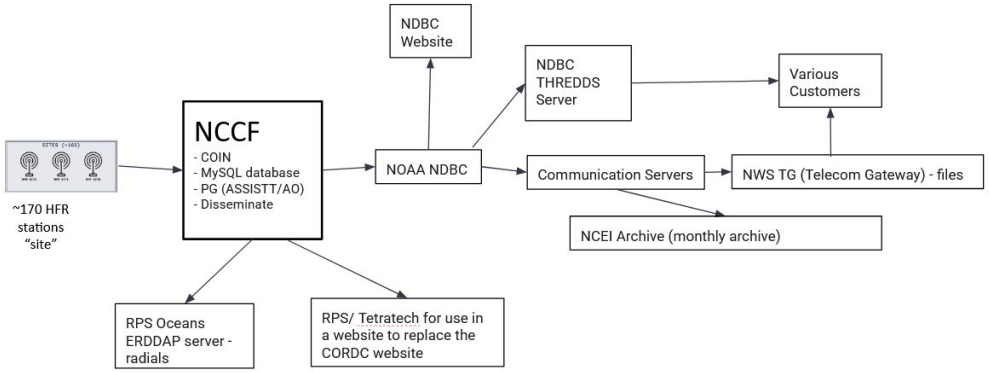
- Secure Ingest
- Science Sandbox
- Dissemination

Mapping NCCF Service Support to HFRNet

Transition to Operations and Beyond

Lessons Learned and Summary

IOOS HFRNet System Diagram (with NCCF) - draft proposed by June 30, 2025



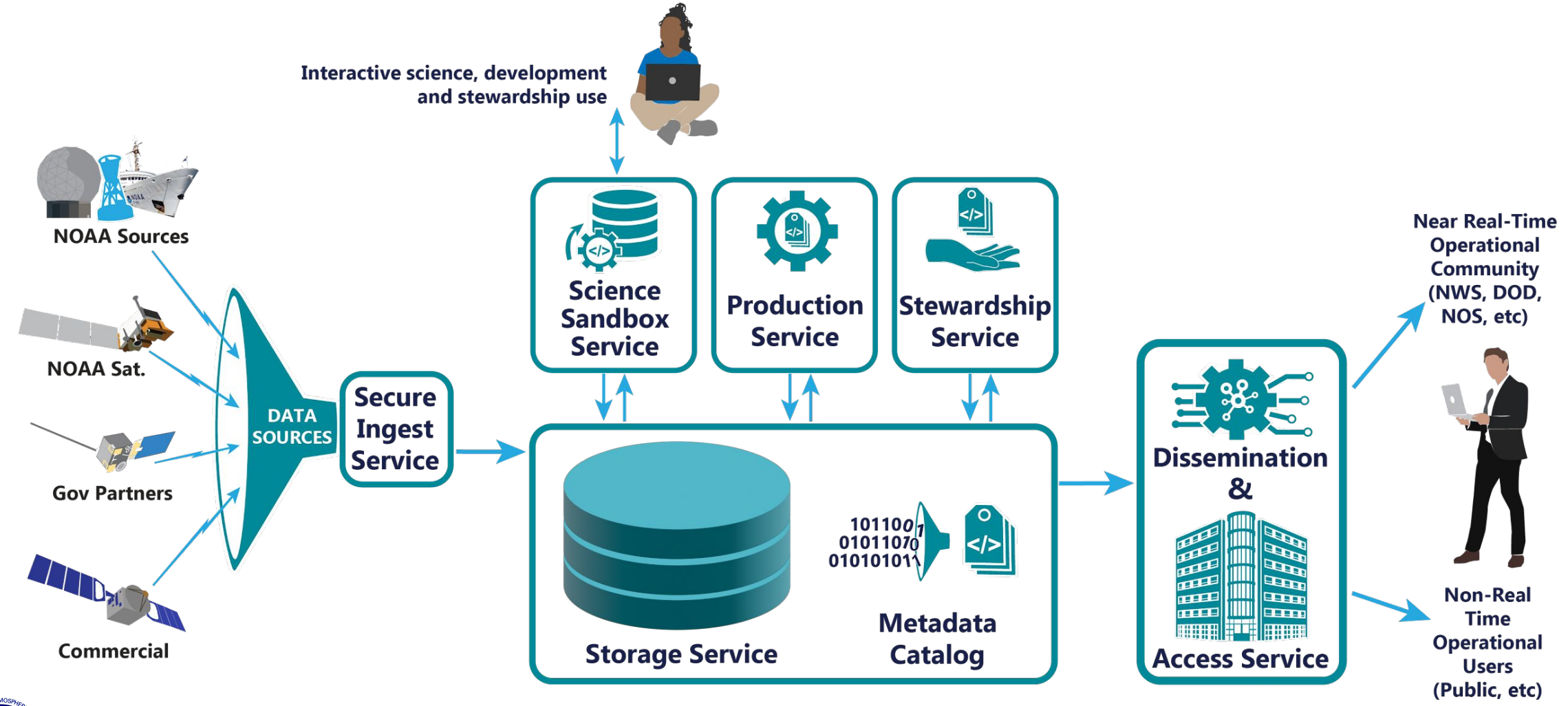
Project Overview

Customer: National Ocean Service - IOOS Office

NCCF Function: Ingest data from ~183 laptops (connected directly or through university run servers), insert data in a SQL database, run data through a Python algorithm (converted from Matlab thanks to the ASSISTT Team), executed using AO (Algorithm Orchestration), inserting the products back into the database and disseminating the data using the new Dissemination-Subscription Fulfillment service to 2 customers (IOOS and NWS/NDBC) pushing data into their S3 buckets

Deadline: Phase 1 is to be completed by June 30, 2025

NESDIS Common Cloud Framework (NCCF)



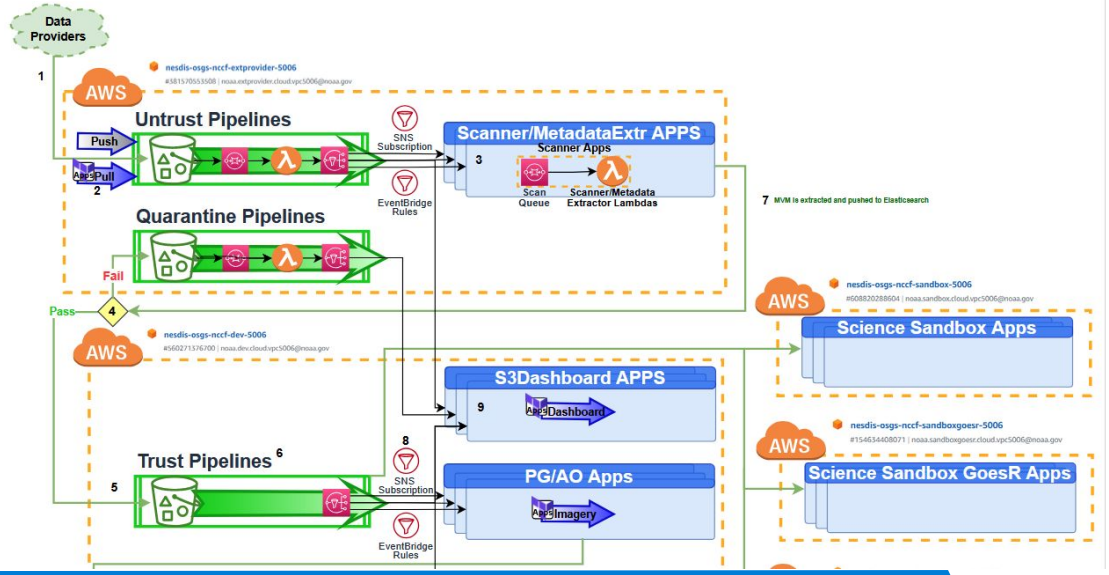
NCCF Operational Capabilities Today

Core Service	Operation Status	Metrics
Secure Ingest	Operational - Q1 FY21	Operational gateway to NCCF. Provides a single secure point of entry for all data types for NESDIS, NOAA and external partners, ensuring all data is safe for use through tailored checks.
Production	Operational - Q3 FY21	Operational fault-tolerant environment for data processing including support for Product Generation with integrated orchestration and processing services.
Science Sandbox	Operational - Q3 FY24	Virtual sandbox environment provisioned for developers, scientists, and collaborators to test algorithms, visualize data, and conduct basic research.
Stewardship	Operational - Q1 FY25	Storage and preservation of data to support other NCCF service, ensuring end-to-end management of data.
Dissemination & Access	Operational Promotion - Q2 FY25*	Provides low-latency data delivery for all NESDIS data within the NCCF.

Secure Ingest

Works with customers to ensure data needs are met either through current NCCF holdings or through ingest process.

- Data ingests variety of formats through one time or subscription model
- Data Push/Pull utilizing HTTPS/FTPS/SFTP/FTP

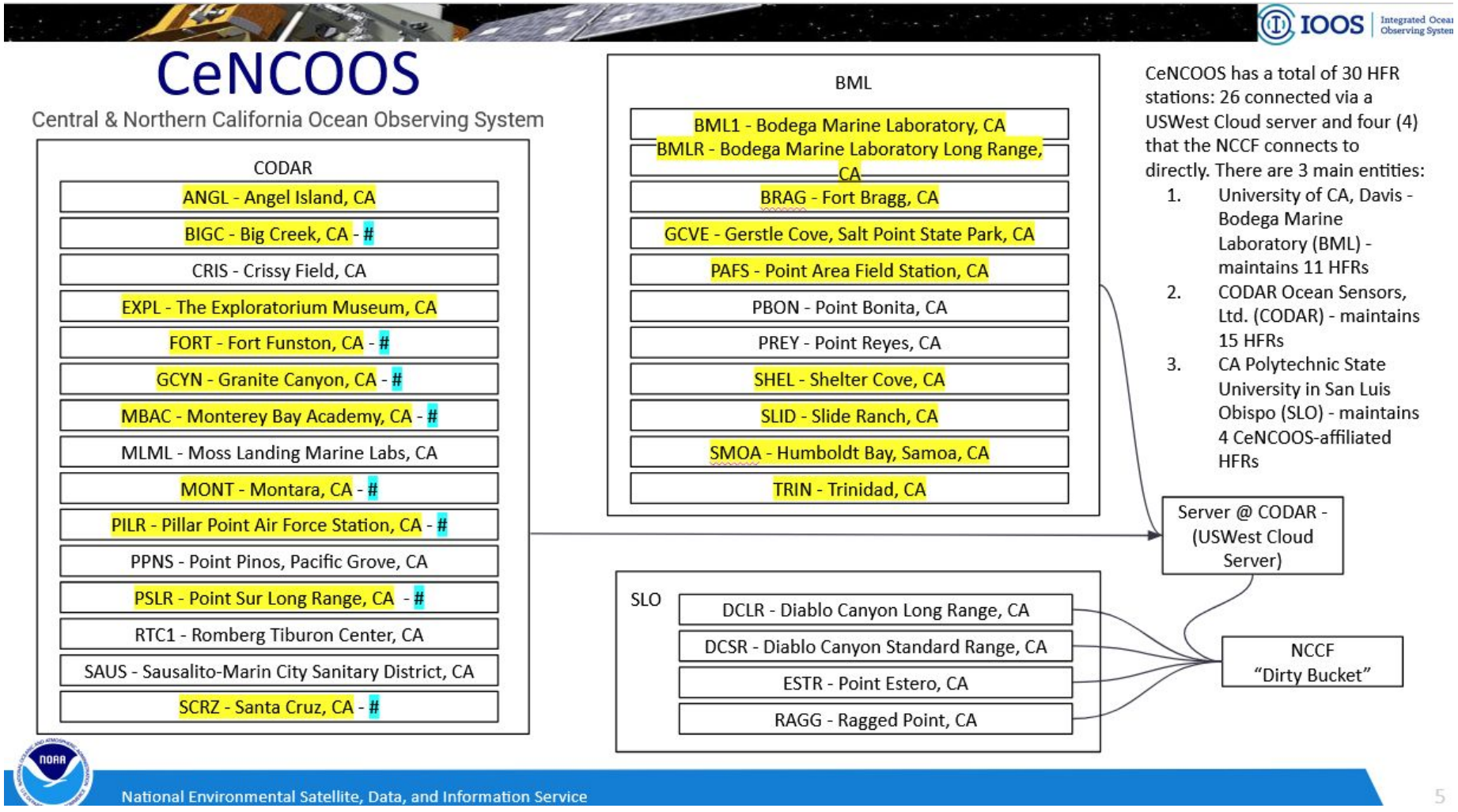


Secure Ingest - HFRNet

Organized data by RA - worked with IOOS PM and station operators to ensure connectivity and troubleshoot issues. Originally identified ~160 or so, current count (without Canadian stations) is 183

he 11 IOOS Regional Associations + Canada		https://ioos.noaa.gov/regions/	Brian provided spreadsheet with connection info	Fishbone Diagram updated	Diwakar verified connectivity?	Data flowing into Trust Bucket? (Puller & Scanner)	Data flowing into Database? (ETL process - Extract Transform Load)	Domain (see below for more info)	Stations with wave data
1	AOOS	Alaska Ocean Observing System	Yes	4	Done	Done		AKNS & GAK	
2	CariCOOS	Caribbean Coastal Ocean Observing System	Yes	8	Done	Done - 6 done.	2 missing due to parts issues (FRDO, THOM)	PRVI	2 - FRDO and SBAY (FRDO - no, down; SBAY - yes)
3	CeNCOOS	Central & Northern California Ocean Observing System	Yes	30	Done	26 done.	4 left - CRIS, ESTR, PBON, PREY	USWC	8 (BIGC, FORT, GCYN, MBAC, MONT, PILR, PSLR, SCRZ)
4	GLOS	Great Lakes Observing System	Yes	2	Done	3/4/2025		GLNA	
5	GCOOS	Gulf of Mexico Coastal Ocean Observing System	Yes	18	In Process	sprint 25-2-5 (3/26/25) - in process - 13 done	5 left - PLA8, CPLF, PINS, SSDE, SWPP	USEGC	
6	MARACOOS	Mid-Atlantic Regional Association Coastal Ocean Observing System	Yes	37	All 37 come from Rutgers Boardwalk server	sprint 25-2-5 (3/26/25) - 33 done	4 left - BRNT, CEDR, MNTK, MVCO	USEGC	6 (BRMR, CMPT, HLGT, SEAB, SILD, SPRK) - all good
7	NERACOOS	Northeastern Regional Association of Coastal Ocean Observing Systems	Yes	10	4 come from Rutgers' Boardwalk server; the other 6 from WHOI	sprint 25-2-6 (4/2/25) - 4 from Rutgers done	6 from WHOI - no connectivity as of 3/18/2025 at 1:00 PM ET per Clement. Waiting on 3/11/2025 response from Anthony.	USEGC	
8	NANOOS	Northwest Association of Networked Ocean Observing Systems	Yes	12	Changing them to SFTP; working with Mike Kosro to test	sprint 25-2-4 (3/12/25) - 0 done	Oregon State needs to convert to SFTP before we can touch them. 3/18/2025 at 1:00 PM ET - Brian Z says Mike Kosro is waiting on input from us. mike.kosro@oregonstate.edu	USWC	
9	PacIOOS	Pacific Islands Ocean Observing System	Yes	7	Done	Done		USHI	
10	SCCOOS	Southern California Coastal Ocean Observing System	Yes	30	In Process - come from 4 different servers; UCSB - done	11 done.	SIO - 10 (no response from Brian Kim - 2/24/2025); CPSU - 5 (connected on 3/18/2025); USC - 6 (connection issues 3/18/2025); UCSB - 9 (good on 3/14/2025)	USEGC	
11	SECOORA	Southeast Coastal Ocean Observing Regional Association	Yes	25	In Process	21 done.	8 left - CNS, FDS, HATY, HTR, JEFF, STF, TRS, VEN	USWC	
12	Canada	HFRs that are part of U.S. IOOS's Canadian counterpart, the Canadian Integrated Ocean Observing System (CIOOS).	Not yet		Moved to phase II			USWC	
			11	183	80.9%	53.0%			16

Fishbone Diagrams

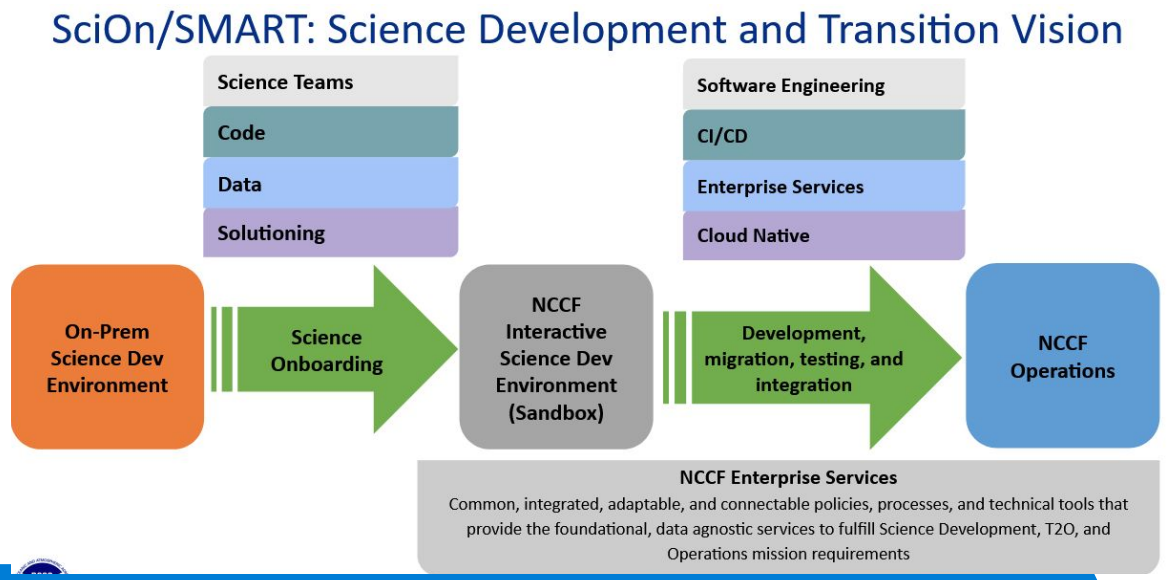


- Illustrates data flows from the laptops/university servers to NCCF
- Important for Security to understand data flow

Science Sandbox Service

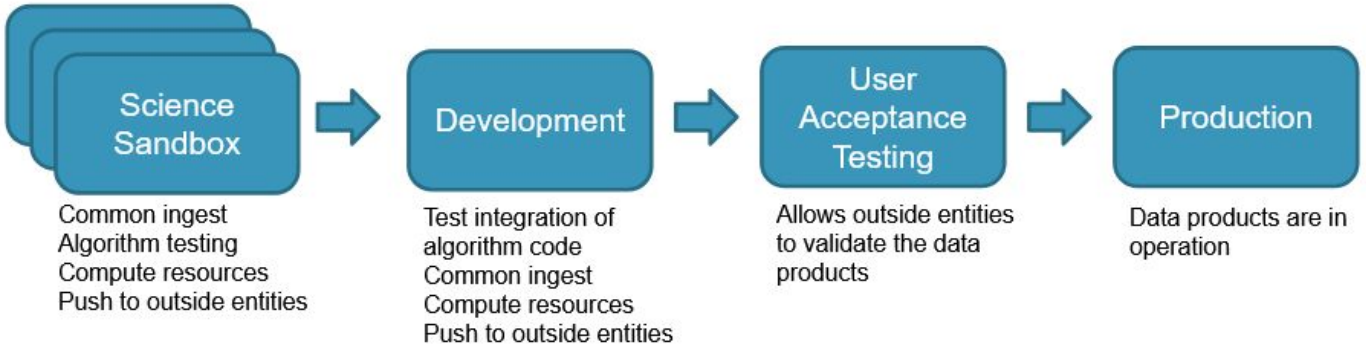
The NCCF Science Sandbox (aka the Interactive Science Development Environment) is:

- A low barrier to entry and flexible environment where day to day science and innovation work can occur
- A common development environment for developers and scientists to develop and test scientific algorithms, visualize data, and conduct basic research
- First stop for Science Teams looking to move their routine product generation work to the NCCF
- Connected to NCCF Common Services
  - Ingest
  - Gitlab CI/CD Pipelines
  - Shared Tools Repository



Science Sandbox Service - HFRNet

From Algorithm to Production: NCCF Environments



With no science team working in the NCCF, the sandbox provides an environment for the algorithm transition team (ASSISTT) and the Application and Dataset Migration Team (ADMT) to incorporate the reconfigured database and code and test these with the live datastreams and Algorithm Orchestration (AO) Service before promoting the application to our 5065 UAT environment.

NCCF Dissemination

Dissemination: Supports delivery of NESDIS products to customers. Current methods of dissemination:

- Legacy Distribution - Push to S3 bucket/FTPS,
  - limited data support
  - no metadata

- Subscription Fulfillment Service - FTPS or SFTP pull, S3 push to customer
  - additional flexibility
  - metadata utilization

Dissemination Service - HFRNet

Deployed to UAT - March 2025

Full deployment to PROD - ~May 2025

Tailored outputs for NDBC and IOOS/TetraTech

Dissemination via Subscription Service

HFRNet NetCDF Total Surface Current Velocity Data Products (available as: hourly, 25-hour average, monthly average, and annual average)				
	500 m	1 km	2 km	6 km
U.S. West Coast (USWC)	x	x	x	x
U.S. East & Gulf Coast (USEGC)		x	x	x
Puerto Rico & Virgin Islands (PRVI)			x	x
U.S. Hawaii (USHI)		x	x	x
Alaska North Slope (AKNS)				x
Gulf of Alaska (GAK)		x	x	x
Great Lakes North America (GLNA)	x	x	x	x

Data Flow Table:

Input: Read in text files (radial and waves) collected by the various HFR stations (ingest and serve back out)
Output: netCDF files (produced from Python code)
Output: the SQL database

#	Data Type	Why does RPS/TT need it?	Where is it coming from?	Where will the data reside?	How can TetraTech get the data from where it resides?
1a	Radials from all HFRs	ERDDAP (RPS Oceans); IOOS.us website	HFR stations	(where will the text files reside once it reads in) - files in native format (unprocessed, serve to TT)	Ex. FTP site S3 bucket transfer/copy (managed by NCCF) Will also need to be eventually provided to NCEI
1b	Wave data where available (from some HFRs)	ERDDAP (RPS Oceans); IOOS.us website	HFR stations	(where will the text files reside once it reads in) - files in native format (unprocessed, serve to TT)	Ex. FTP site S3 bucket transfer/copy (managed by NCCF) Will also need to be eventually provided to NCEI
2	TVs (Total Velocity Vectors) • Grid resolutions: 500 m, 1 km, 2 km, 6 km • Those resolutions are available at temporal periods: Hourly & averages (25 hr, month, and year)	IOOS.us website: Display vectors and time-series plots.	Produced by the NCCF		(needed by RPS/TT, NDBC and eventually NCEI) Public S3 bucket or sync to RPS S3 mirror
3	TV data access to 2012 (for RPS to make viewable on a map on the website)	IOOS.us website	These data are currently at SIO. Who will import and serve this archive… NCEI?, NCCF dissemination?, NDBC?, RPS/TT?		
4	Sensor information/diagnostics (from the SQL database diagramed at hfradar_eer.pdf) *-need specifics on this*	IOOS.us website	Exists now in SQL database; how to access these data from the NCCF?	This is the 2 SQL databases that the Python code will be updating	Database replication? Read-only access, API hosted by NCCF, or replicate database to RPS
5	Database - based on files in headers in raw files				

Mapping NCCF Service Support to HFRNet

NCCF Service	HFRNet Project Applicability	Improvements over existing HFRNet
Infrastructure	Connects to each laptop/University server	Streamlined access versus portal/node structure
Secure Ingest	Reads in data from each laptop/server every 10 minutes - also reads metadata	Dashboards show status
Database	MySQL database implemented to store data after read into the appropriate location	3 different databases merged into one (1)
Algorithm Orchestration	New version of Product Generation process that runs program to read in live data and generates various time-dimensional products, which are then put back in the database	ASSISTT Team rewrote processing code program from MATLAB to Python
Dissemination Service	Uses new Subscription Fulfillment capability to send data via S3 buckets to both RPS/TT and NDBC.	Dissemination using SF - distributes all results to both RPS/TT and NDBC. Future: Archive; Canadian stations.
Security	Verifies that code passes security scans and any vulnerabilities are addressed	All data collected via SFTP; old system included FTP
OSPO transition	Project Monitoring; Modify project on a 24x7x365 basis	Limited project monitoring; Current system monitored as needed

Transition to Operations and Beyond

So what happens after June 30, 2025?

- OSPO will be monitoring the system and data flows and product quality on a 24x7x365 basis.
- If any data from a station is delayed for 8 hours and then 24 hours, then an email will be sent to operator's group and IOOS PM
- Work with IOOS to add and remove stations and/or change a station's beampattern.

- Future:
  - Move the Archive process into the NCCF instead of via NDBC
  - Add in Canadian stations

Lessons Learned and Summary

Enabled by Stakeholder Engagement

01 Communication is Key

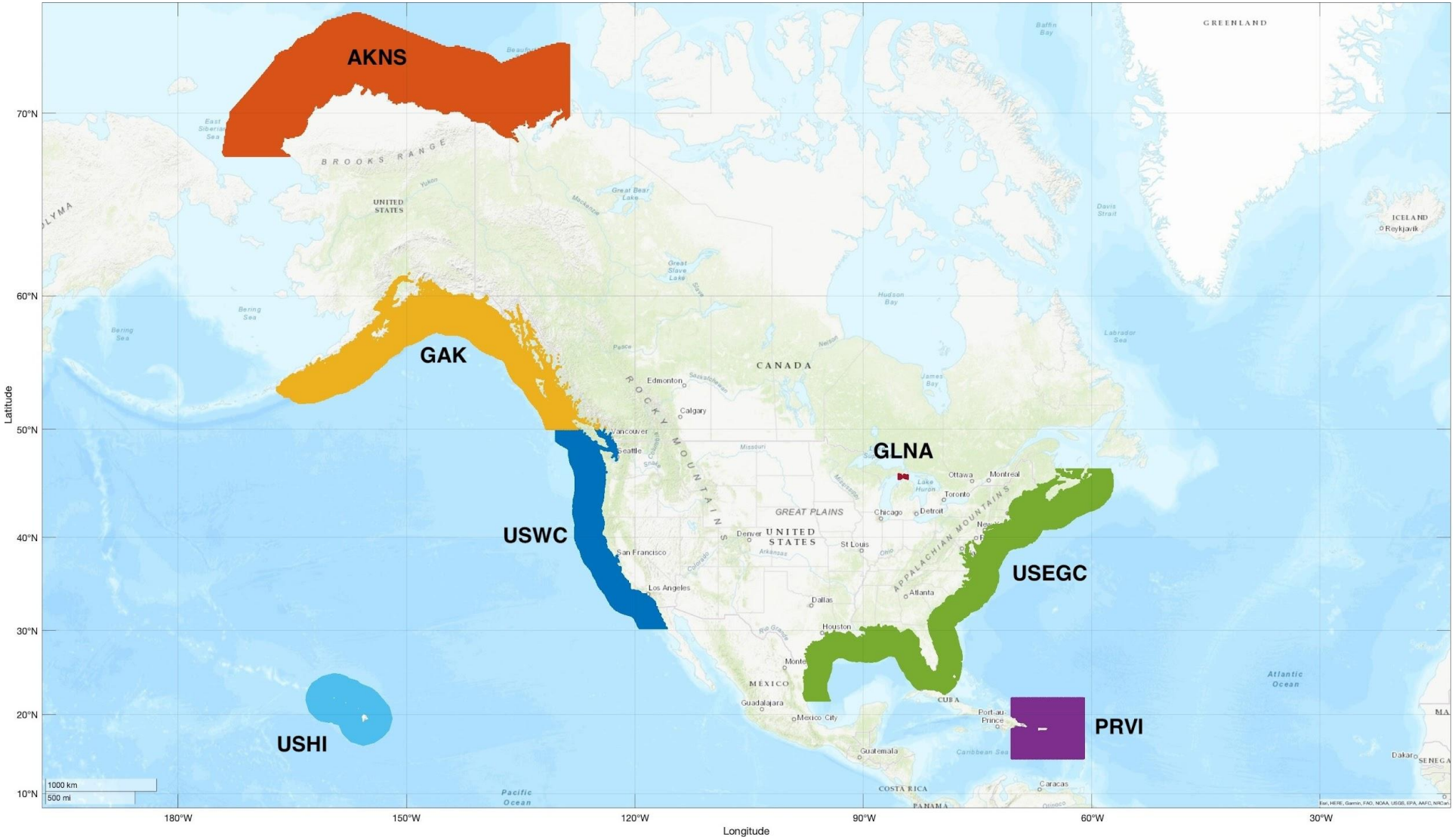
02 Early identification and coordination with stakeholders saves time later on (such as with security process)

03 Being adaptable and responsive is key! Project served as test case for several services to ensure they would be available on our timeline

04 Establishing deliverables and delineating between “nice to have” and “required” is a must for tight deadlines

05 Coordinating across 15+ groups of stakeholders is hard! Regular meeting cad plus multiple communication channels (chat space, document repository, etc..) cuts down on miscommunications and dropped details

Questions?


AKNS
GAK
USWC
GLNA
USEGC
USHI
PRVI
70°N
60°N
50°N
40°N
30°N
20°N
10°N
Latitude
180°W
150°W
120°W
90°W
60°W
30°W
Longitude
1000 km
500 mi
GREENLAND
ICELAND
CANADA
UNITED STATES
MÉXICO
CUBA
COSTA RICA
PANAMA
BROOKS RANGE
ROCKY MOUNTAINS
GREAT PLAINS
APPALACHIAN MOUNTAINS
Edmonton
Calgary
Vancouver
Seattle
San Francisco
Los Angeles
Denver
St Louis
Dallas
Houston
Atlanta
Chicago
Detroit
Toronto
Ottawa
Montreal
Guadalajara
Mexico City
Guatemala
Port-au-Prince
Caracas
Dakar
Reykjavik
Pacific Ocean
Atlantic Ocean
Hudson Bay
Labrador Sea
Davis Strait
Baffin Bay
Great Bear Lake
Great Slave Lake
Bering Sea
East Siberian Sea
Caribbean Sea

Thank you!

For more info, please contact:

Inger Kittle, OCS Product Implementation Branch **inger.kittle@noaa.gov**

Backup Slides

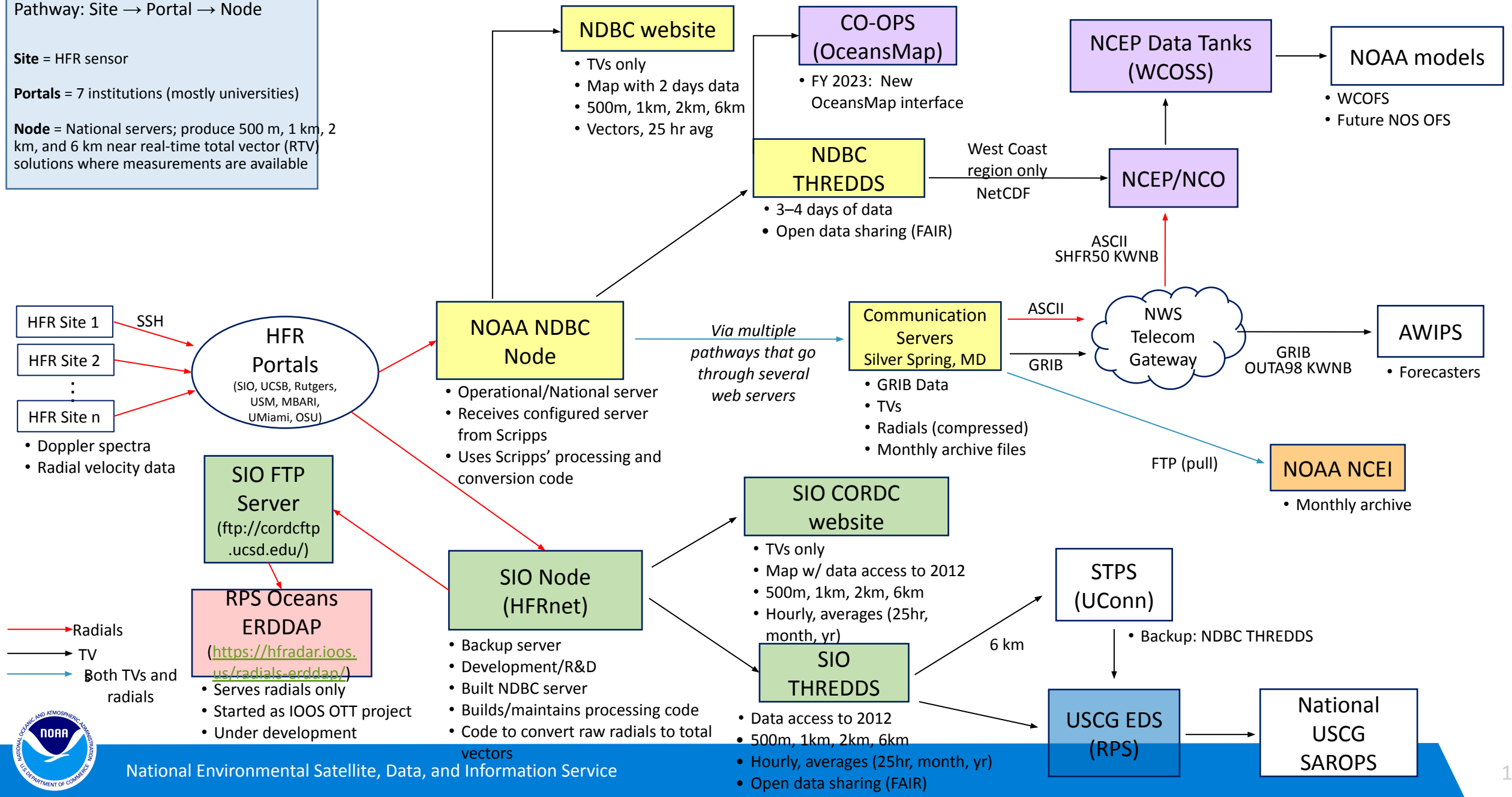
HFRNet: At it is now

Pathway: Site → Portal → Node

Site = HFR sensor

Portals = 7 institutions (mostly universities)

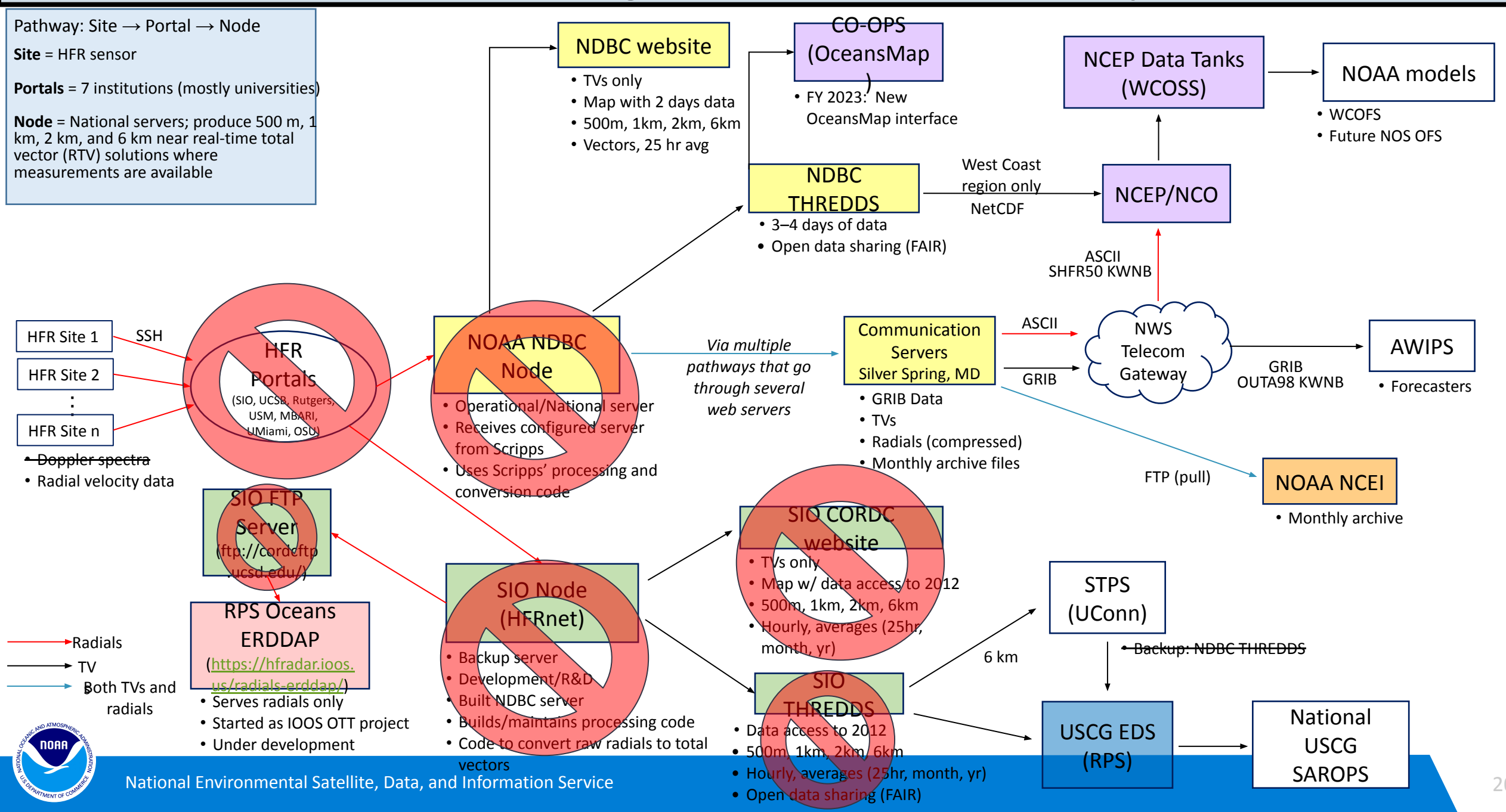
Node = National servers; produce 500 m, 1 km, 2 km, and 6 km near real-time total vector (RTV) solutions where measurements are available



HFRNet: What is being removed with SIO-CORDC's departure

Pathway: Site → Portal → Node

Site = HFR sensor

Portals = 7 institutions (mostly universities)

Node = National servers; produce 500 m, 1 km, 2 km, and 6 km near real-time total vector (RTV) solutions where measurements are available