

OOIFB / DDCI 2018
Technologies and Approaches

Cyberinfrastructure for IOOS and Beyond

Kyle Wilcox
Axiom Data Science
10/30/2018

Company Background



- Founded in 2006
- 20 employees
- Offices in AK, OR and RI

- Mission driven: Build capabilities which accelerate the synthesis and re-use of earth science data

- Federal Partners: NSF, NOAA, USGS, BOEM, NASA, FWS, DARPA, ONR
- Other Partners: MBARI, Shell, UAF, UAA, Carnival

IOOS In a Nutshell



11 regional coastal observing systems comprise the NOAA led U.S. IOOS®

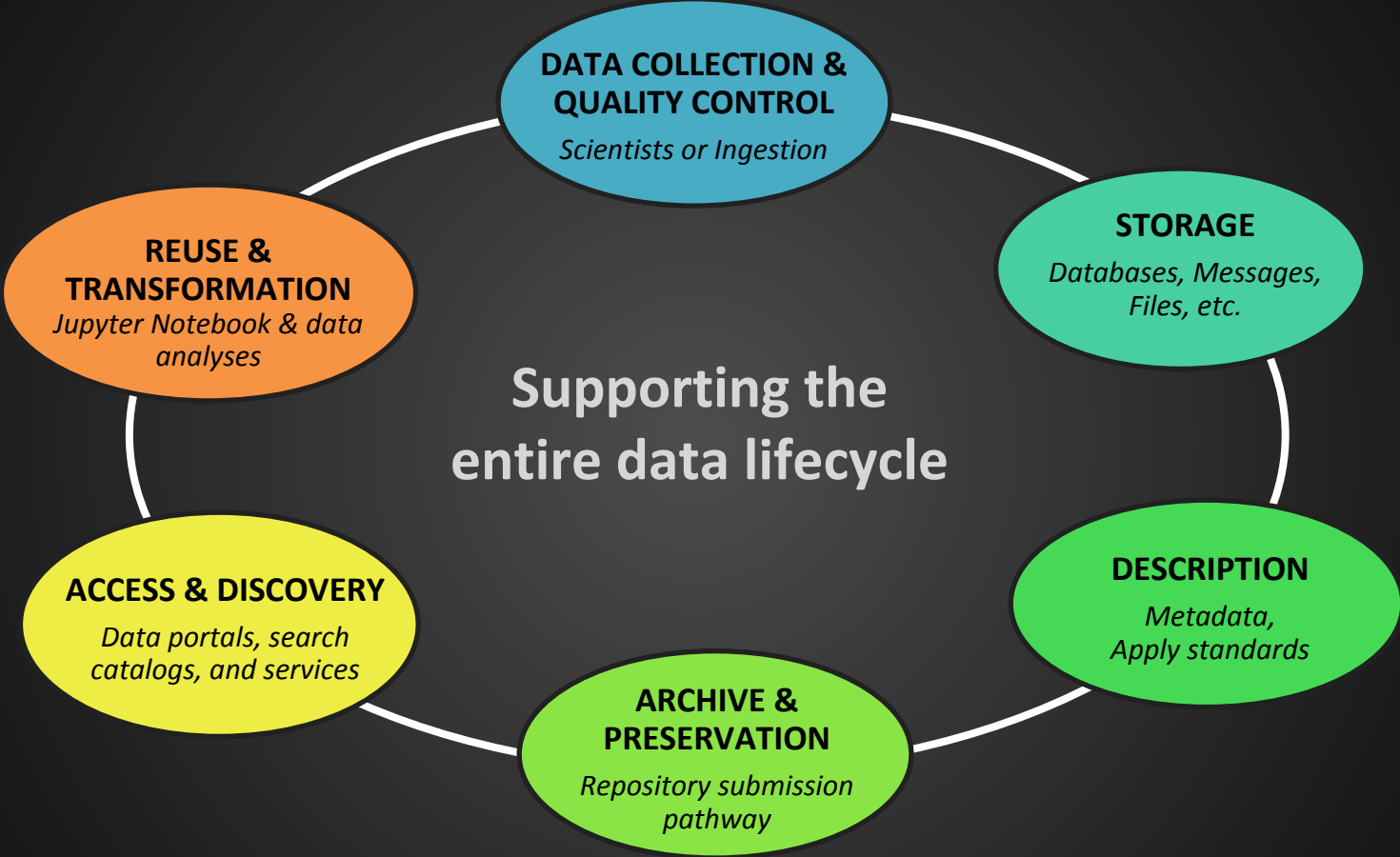
Cyberinfrastructure

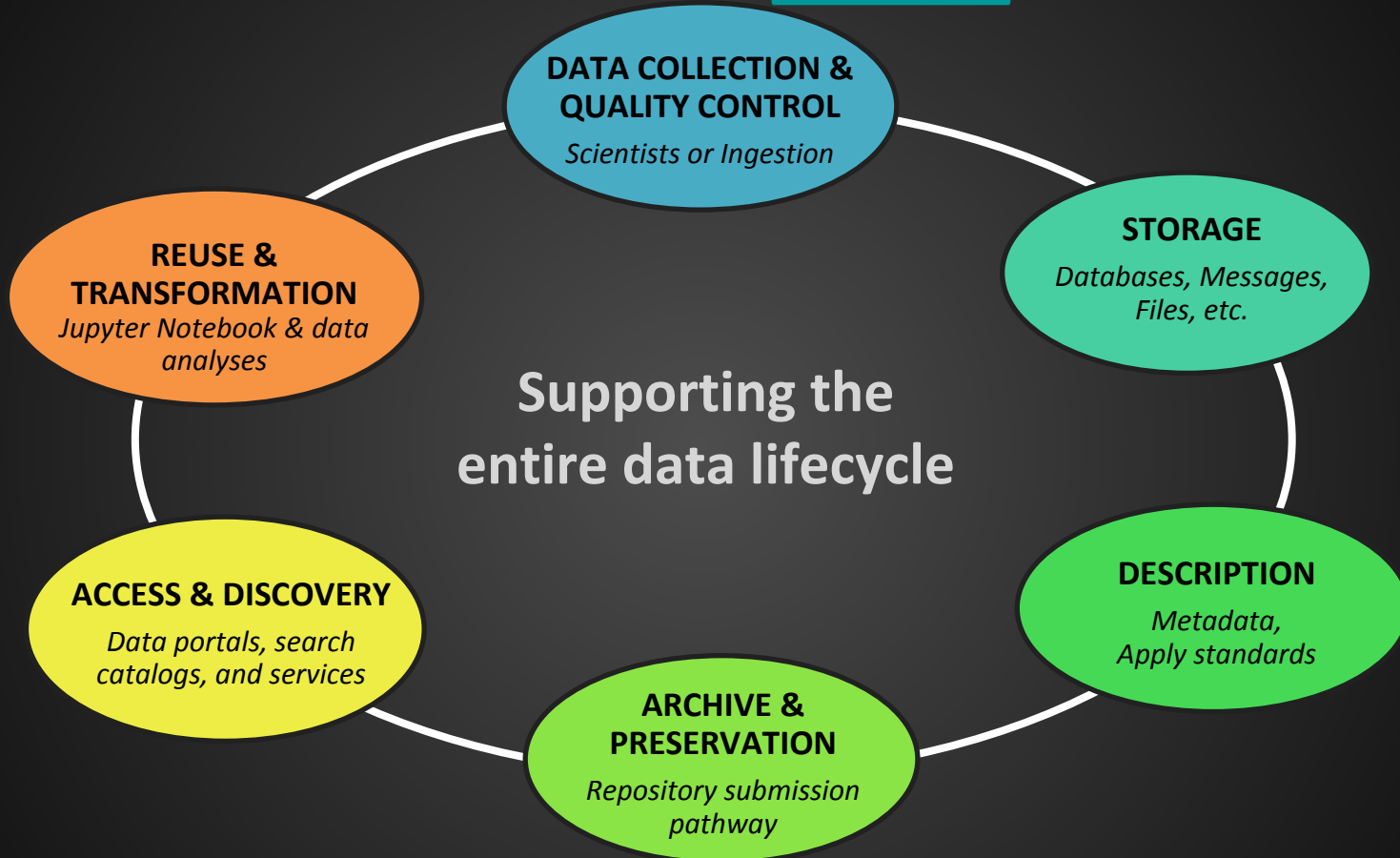
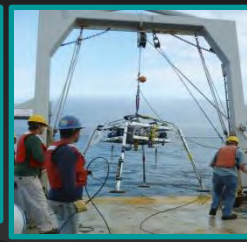
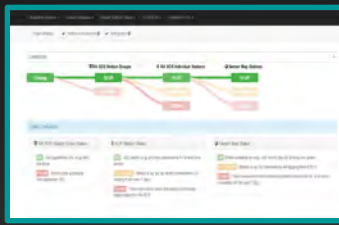
The National Science Foundation defines cyberinfrastructure as:

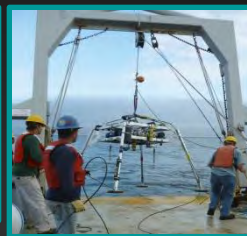
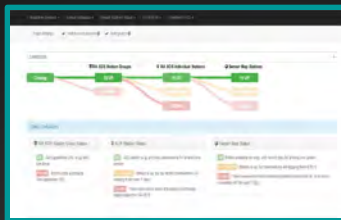
In scientific usage, cyberinfrastructure is a technological and sociological solution to the problem of efficiently connecting laboratories, data, computers, and people with the goal of enabling derivation of novel scientific theories and knowledge.

- *Community developed software, standards and protocols*
- *Scalable compute and storage infrastructure (HPC)*
- *Human capacity - data scientists, data librarians, data coordinators, software engineers*

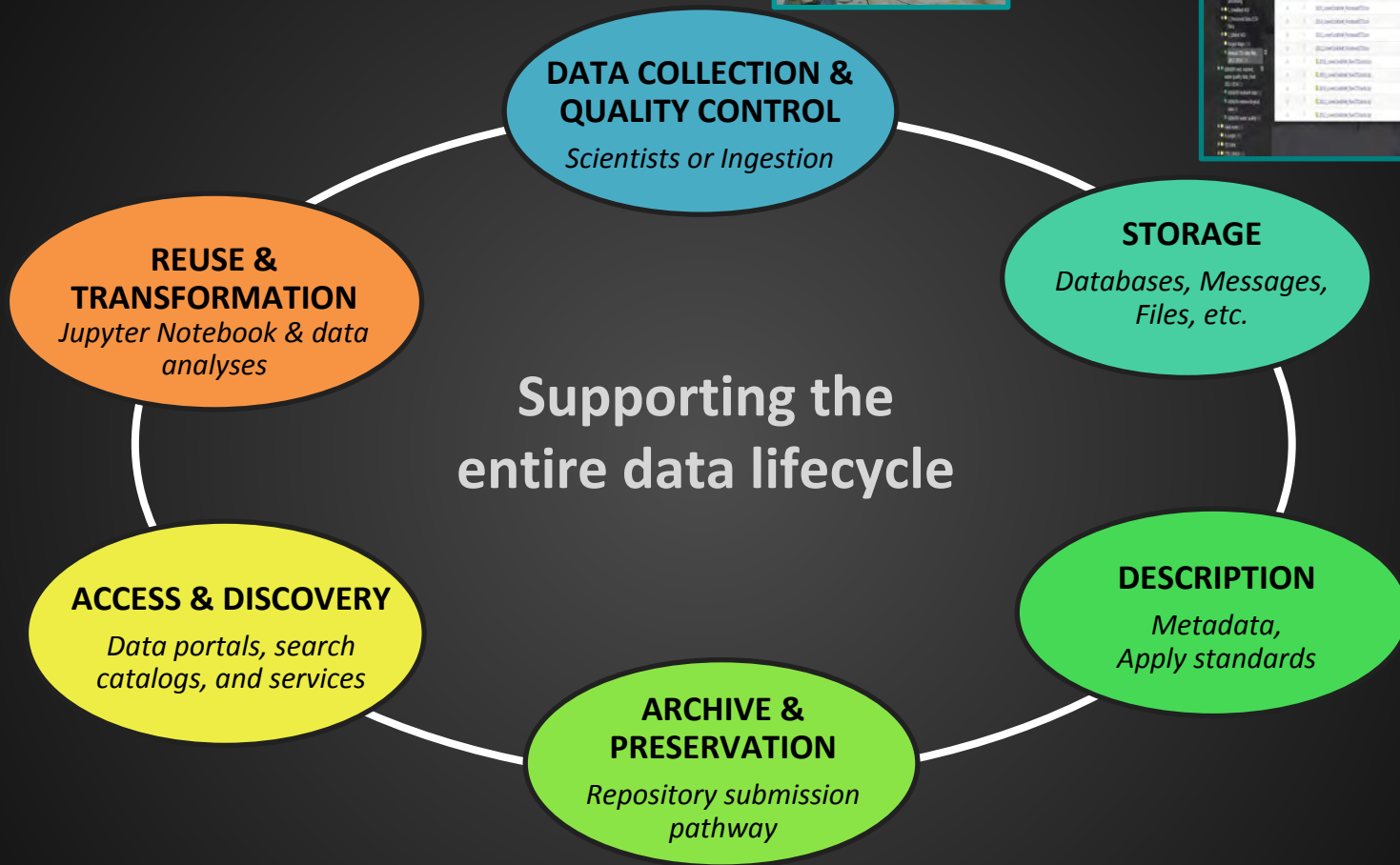
Data Management Lifecycle

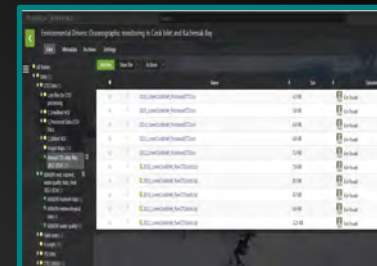
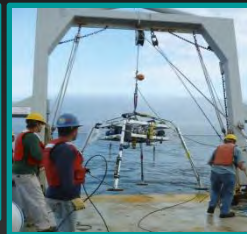
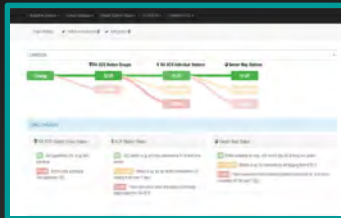






Name	Size	Created
000_LowLatitude_OceanTemp	2100	2018-01-01
000_LowLatitude_OceanSal	2100	2018-01-01
000_LowLatitude_OceanDepth	2100	2018-01-01
000_LowLatitude_SatelliteTemp	2100	2018-01-01
000_LowLatitude_SatelliteSal	2100	2018-01-01
000_LowLatitude_SatelliteDepth	2100	2018-01-01
000_LowLatitude_SatelliteWind	2100	2018-01-01
000_LowLatitude_SatelliteWave	2100	2018-01-01
000_LowLatitude_SatelliteCloud	2100	2018-01-01
000_LowLatitude_SatelliteRain	2100	2018-01-01
000_LowLatitude_SatelliteSnow	2100	2018-01-01
000_LowLatitude_SatelliteIce	2100	2018-01-01
000_LowLatitude_SatelliteSoil	2100	2018-01-01
000_LowLatitude_SatelliteVegetation	2100	2018-01-01
000_LowLatitude_SatelliteLand	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevel	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelChange	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressure	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureChange	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRate	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateChange	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateRate	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateRateChange	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateRateRate	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateRateRateChange	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateRateRateRate	2100	2018-01-01
000_LowLatitude_SatelliteSeaLevelPressureRateRateRateRateChange	2100	2018-01-01





DATA COLLECTION & QUALITY CONTROL
Scientists or Ingestion

STORAGE
Databases, Messages, Files, etc.

REUSE & TRANSFORMATION
Jupyter Notebook & data analyses

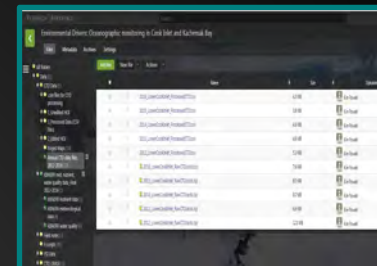
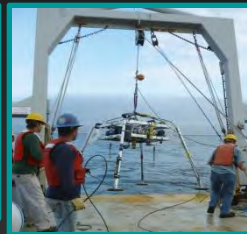
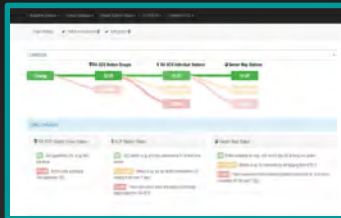
Supporting the entire data lifecycle

DESCRIPTION
Metadata, Apply standards

ACCESS & DISCOVERY
Data portals, search catalogs, and services

ARCHIVE & PRESERVATION
Repository submission pathway





DATA COLLECTION & QUALITY CONTROL
Scientists or Ingestion

STORAGE
Databases, Messages, Files, etc.

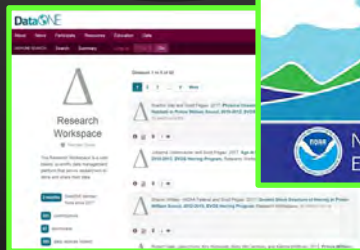
REUSE & TRANSFORMATION
Jupyter Notebook & data analyses

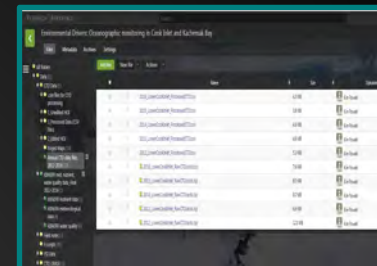
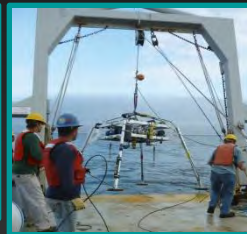
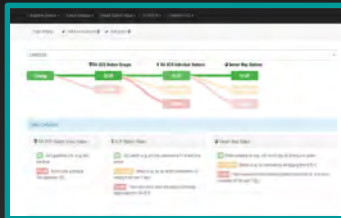
Supporting the entire data lifecycle

DESCRIPTION
Metadata, Apply standards

ACCESS & DISCOVERY
Data portals, search catalogs, and services

ARCHIVE & PRESERVATION
Repository submission pathway





DATA COLLECTION & QUALITY CONTROL
Scientists or Ingestion

STORAGE
Databases, Messages, Files, etc.

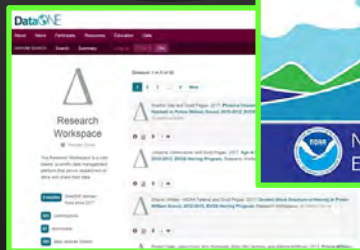
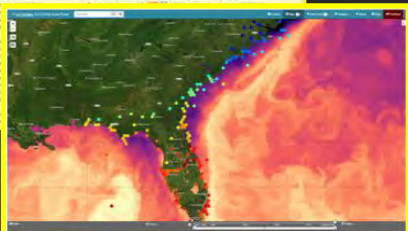
REUSE & TRANSFORMATION
Jupyter Notebook & data analyses

Supporting the entire data lifecycle

DESCRIPTION
Metadata, Apply standards

ACCESS & DISCOVERY
Data portals, search catalogs, and services

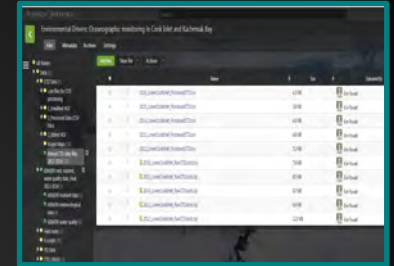
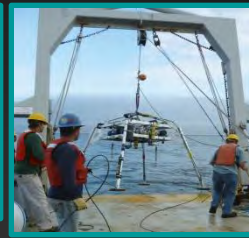
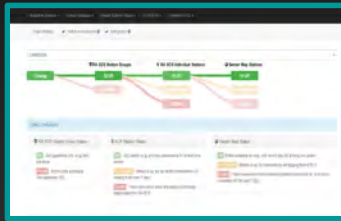
ARCHIVE & PRESERVATION
Repository submission pathway



Supporting the entire data lifecycle

DATA COLLECTION & QUALITY CONTROL

Scientists or Ingestion



Name	Size	Status
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	2.0 GB	Upload

STORAGE

Databases, Messages, Files, etc.

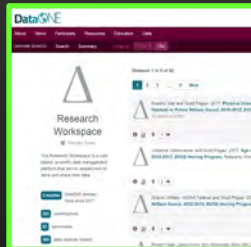
DESCRIPTION

Metadata, Apply standards



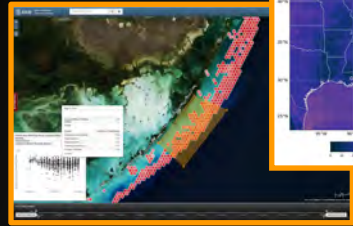
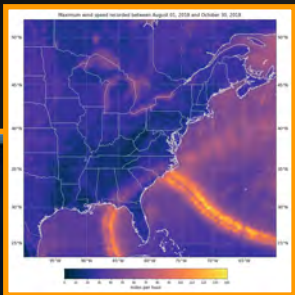
ARCHIVE & PRESERVATION

Repository submission pathway



REUSE & TRANSFORMATION

Jupyter Notebook & data analyses

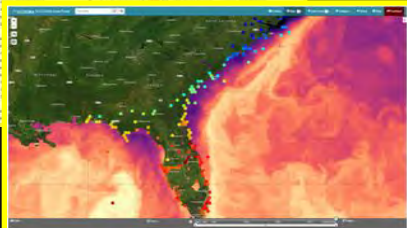


ACCESS & DISCOVERY

Data portals, search catalogs, and services



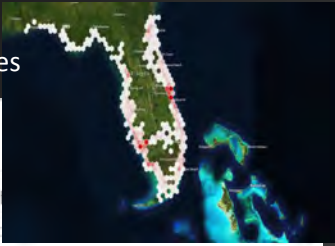
Dataset Name	Access URL	Access Method	Access Type
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public
2023 Environmental Oceanographic Monitoring in Cook Inlet and Kachemak Bay	https://aodn.cerl.noaa.gov/erddap/	HTTP	Public



IOOS DMAC Approach

Biodiversity

count, richness, diversity indices



Platforms

moorings, shore stations



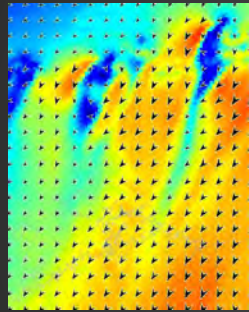
Products

skill assessment, shoreline change, etc.



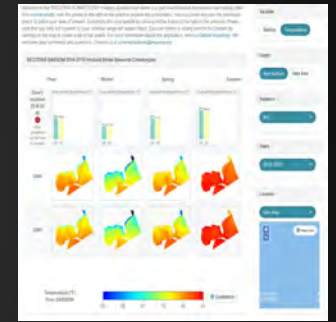
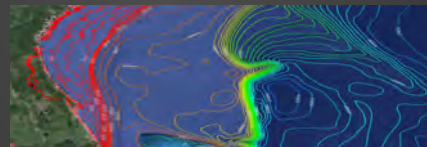
Grids

models, satellite, radar

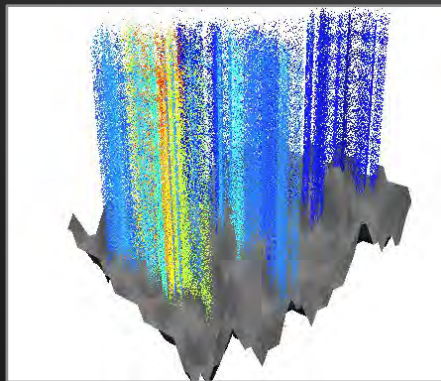


GIS

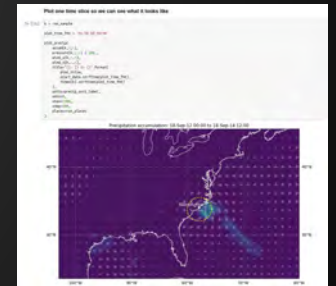
Habitat types, bathymetry, fishing zones, etc.



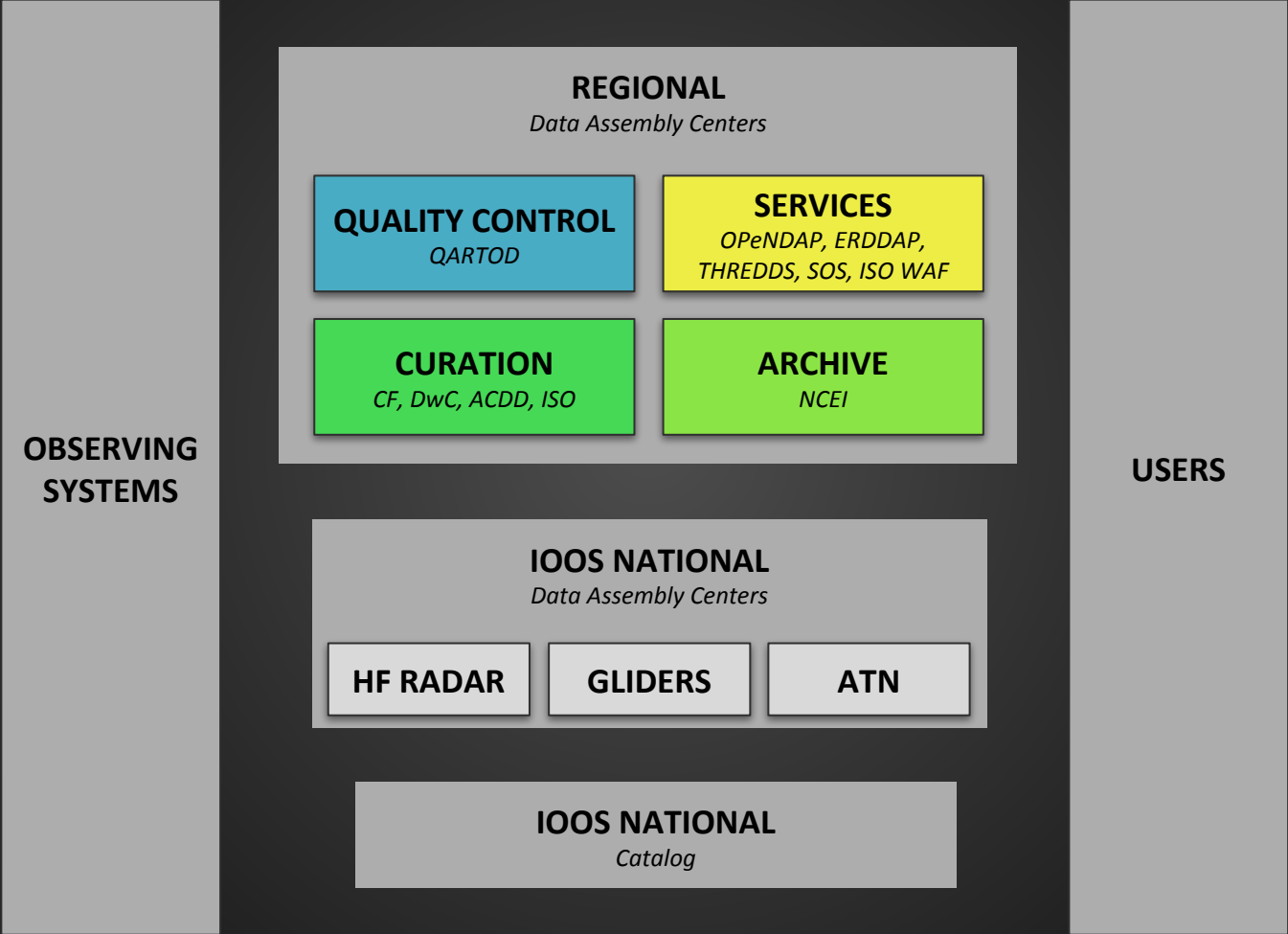
Moving Platforms



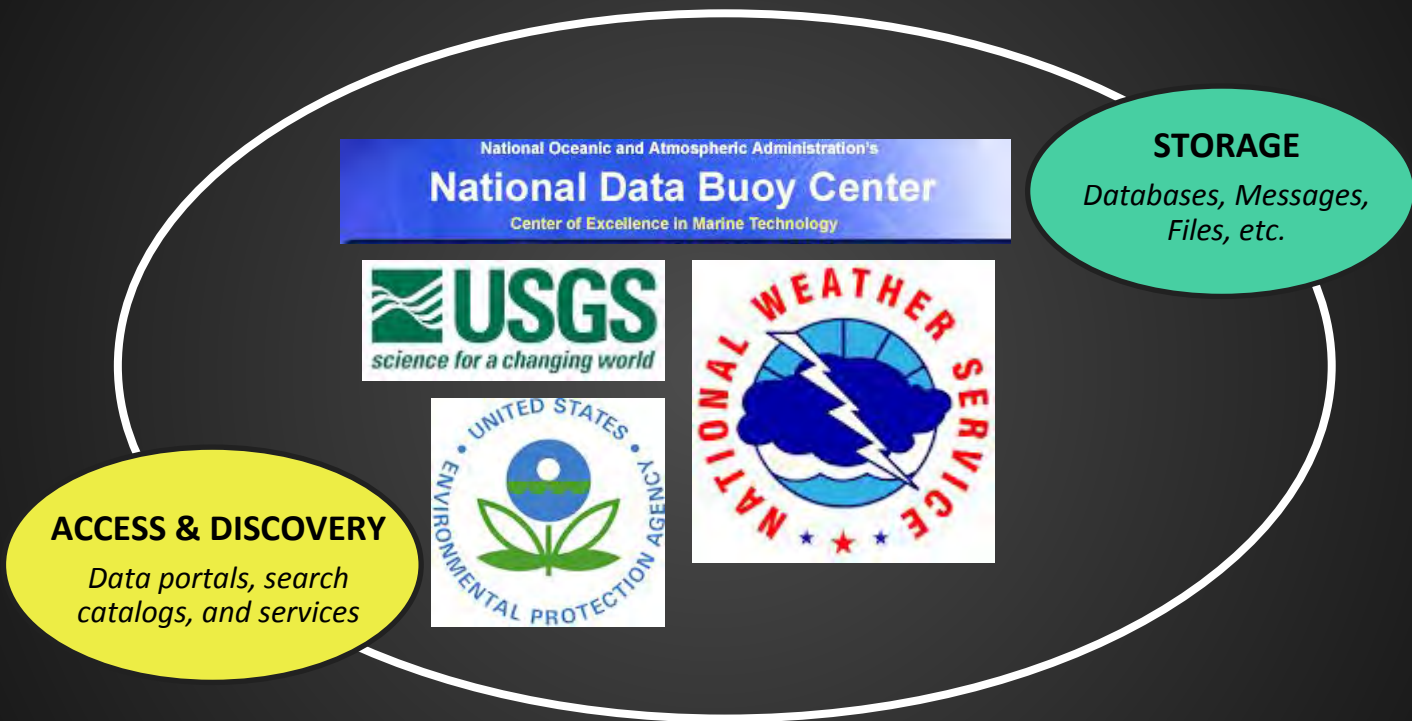
Unstructured Data



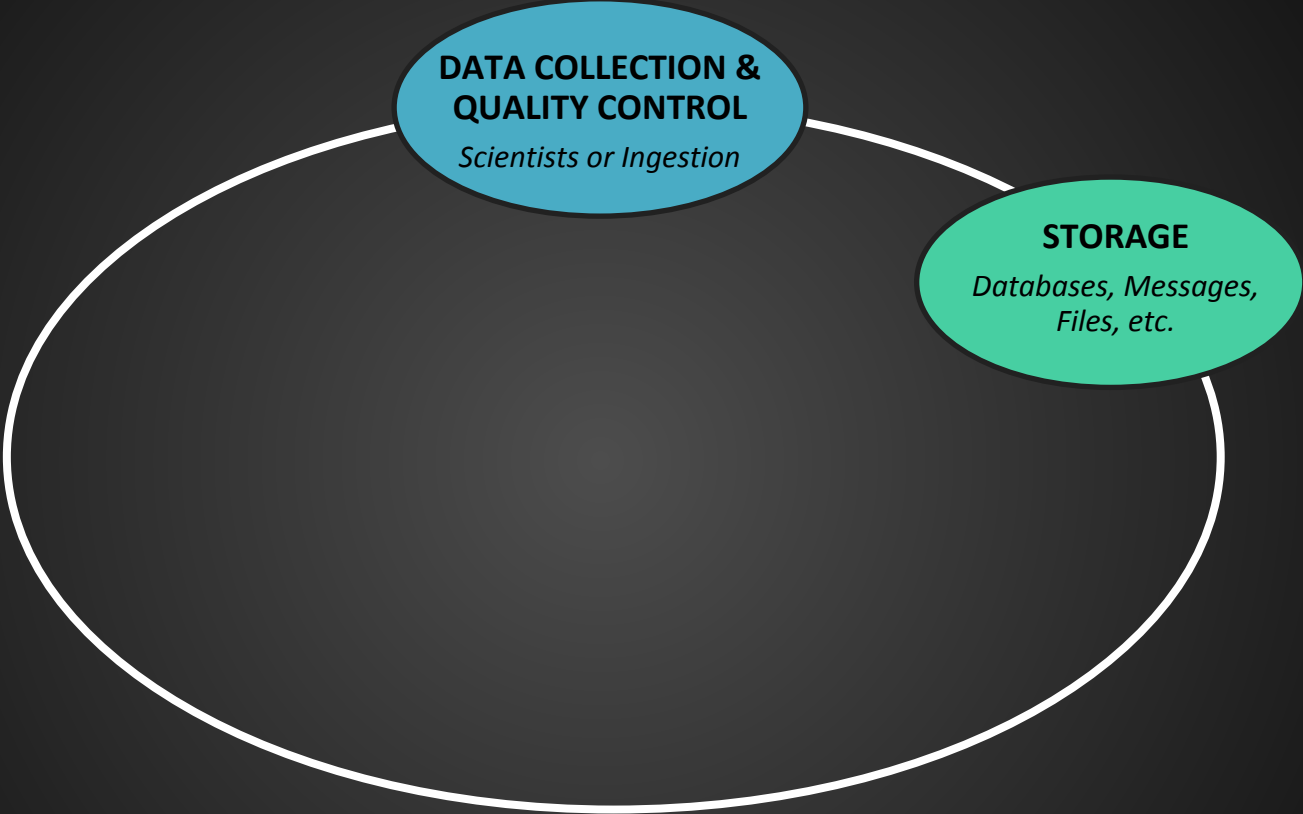
IOOS DMAC Approach



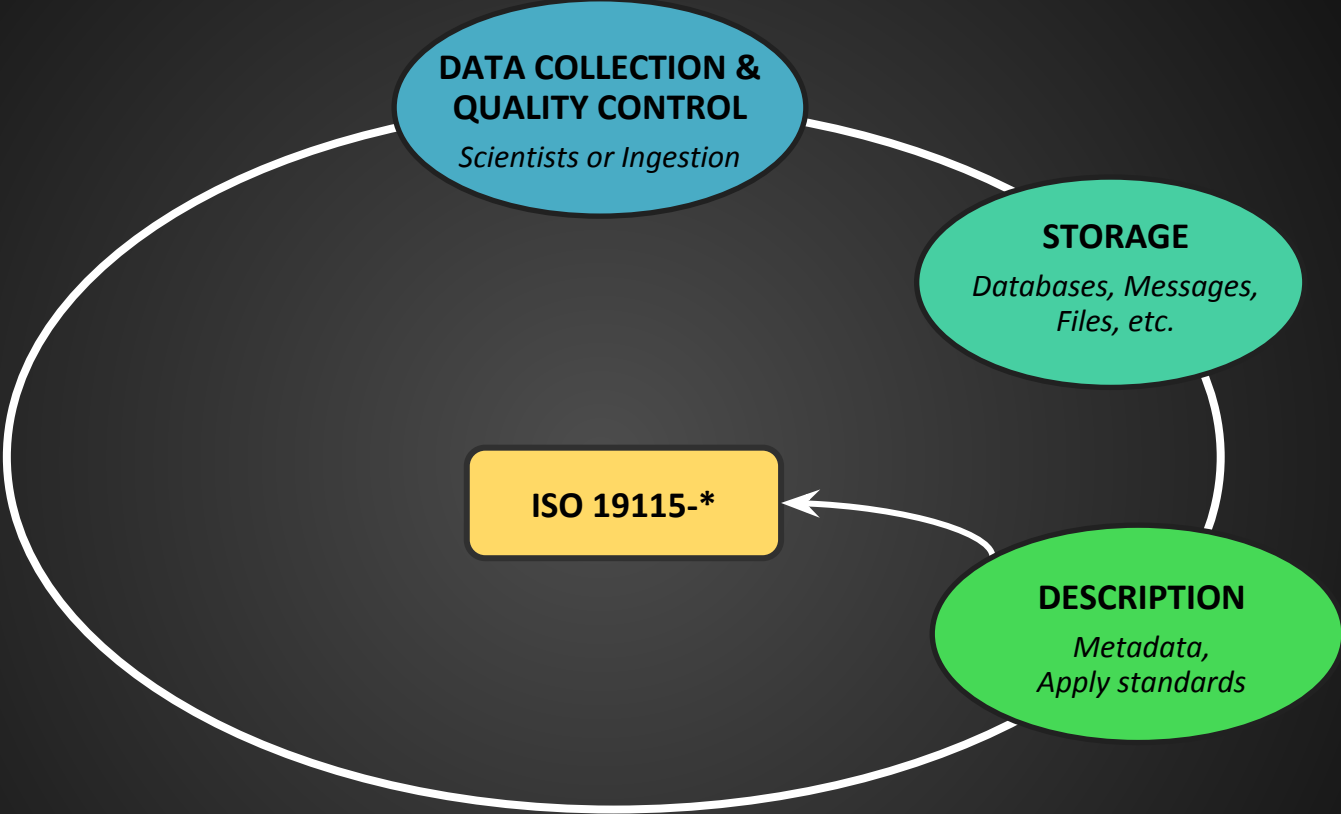
Funded vs. Unfunded



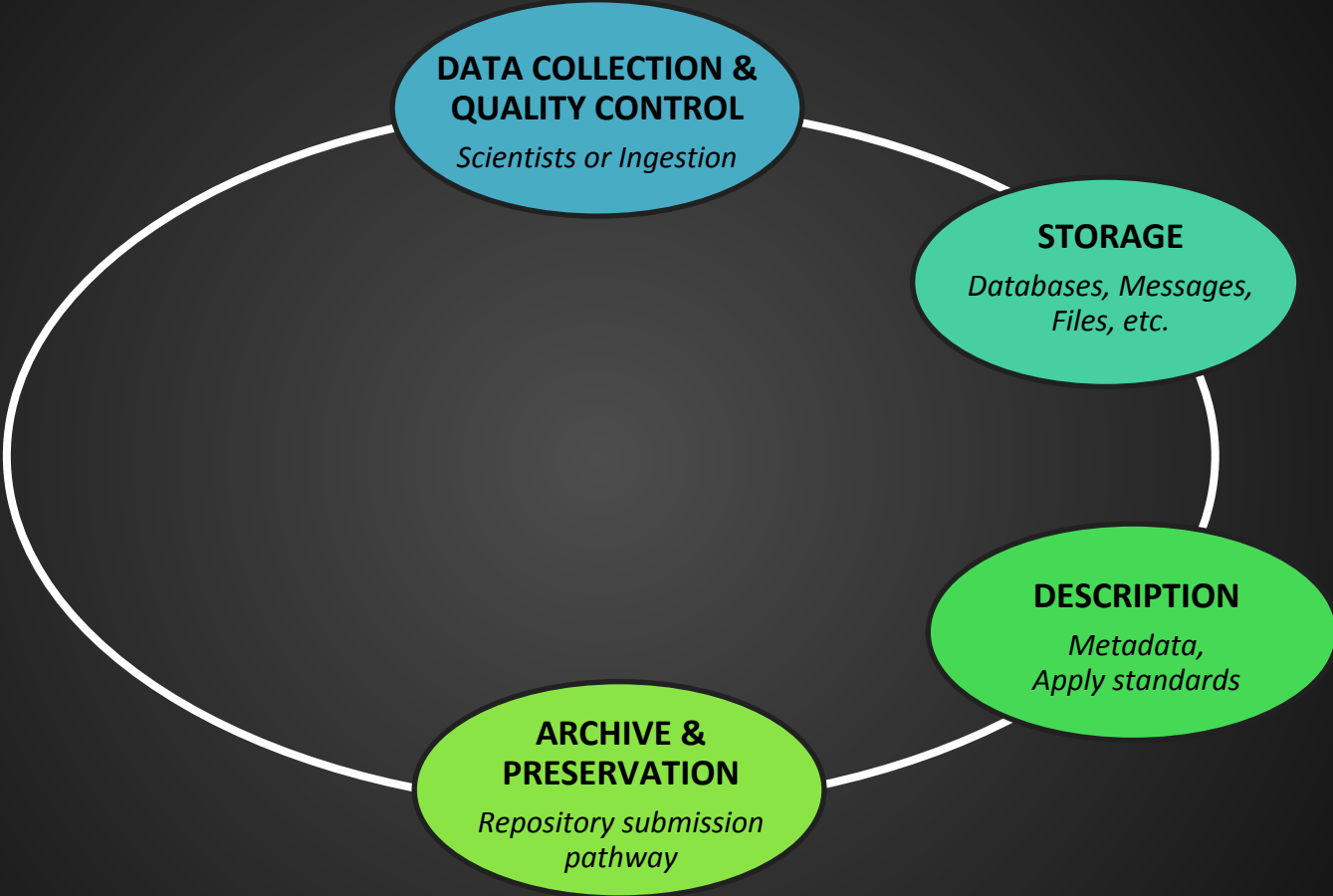
DAC Dataset Flow



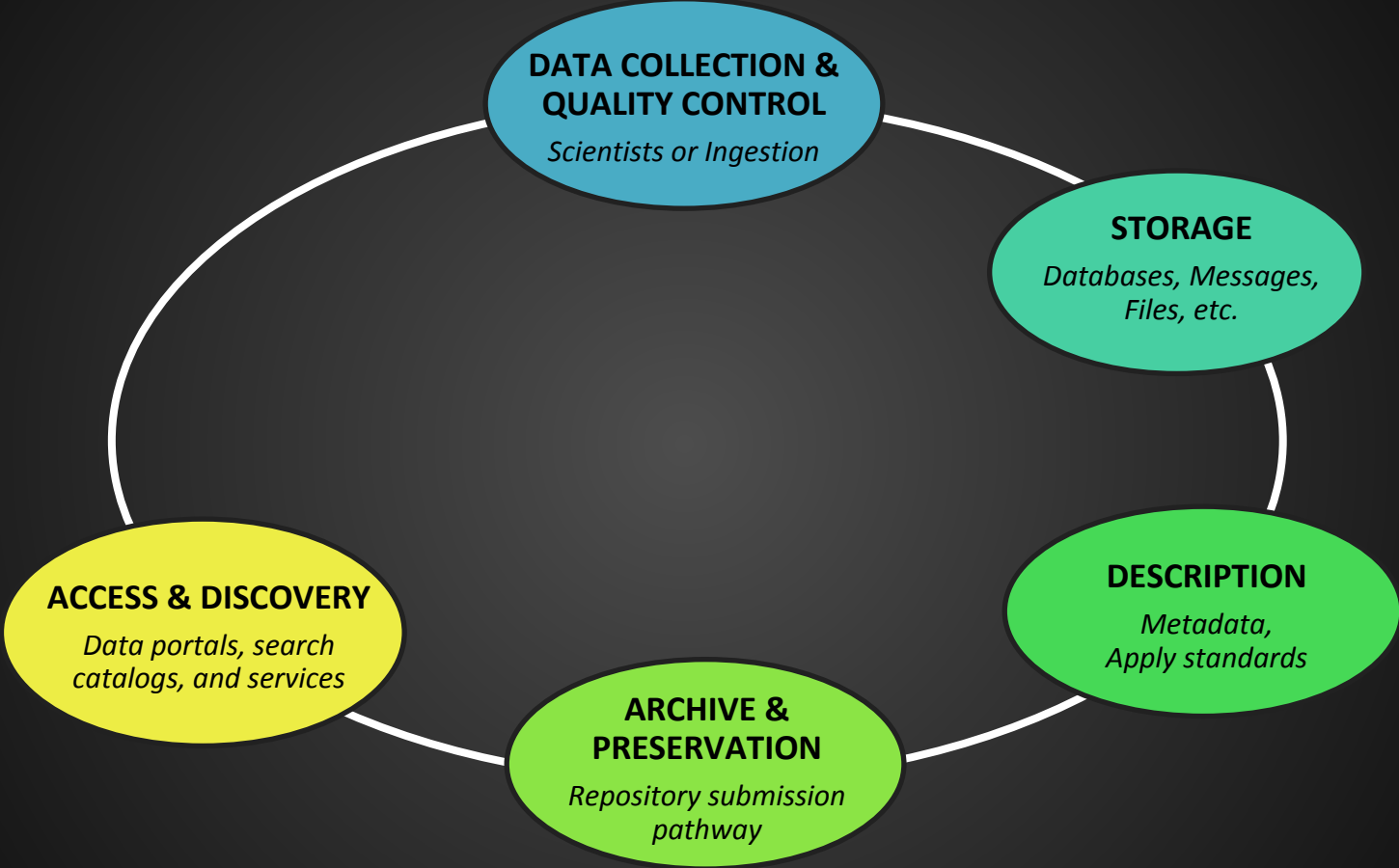
DAC Dataset Flow



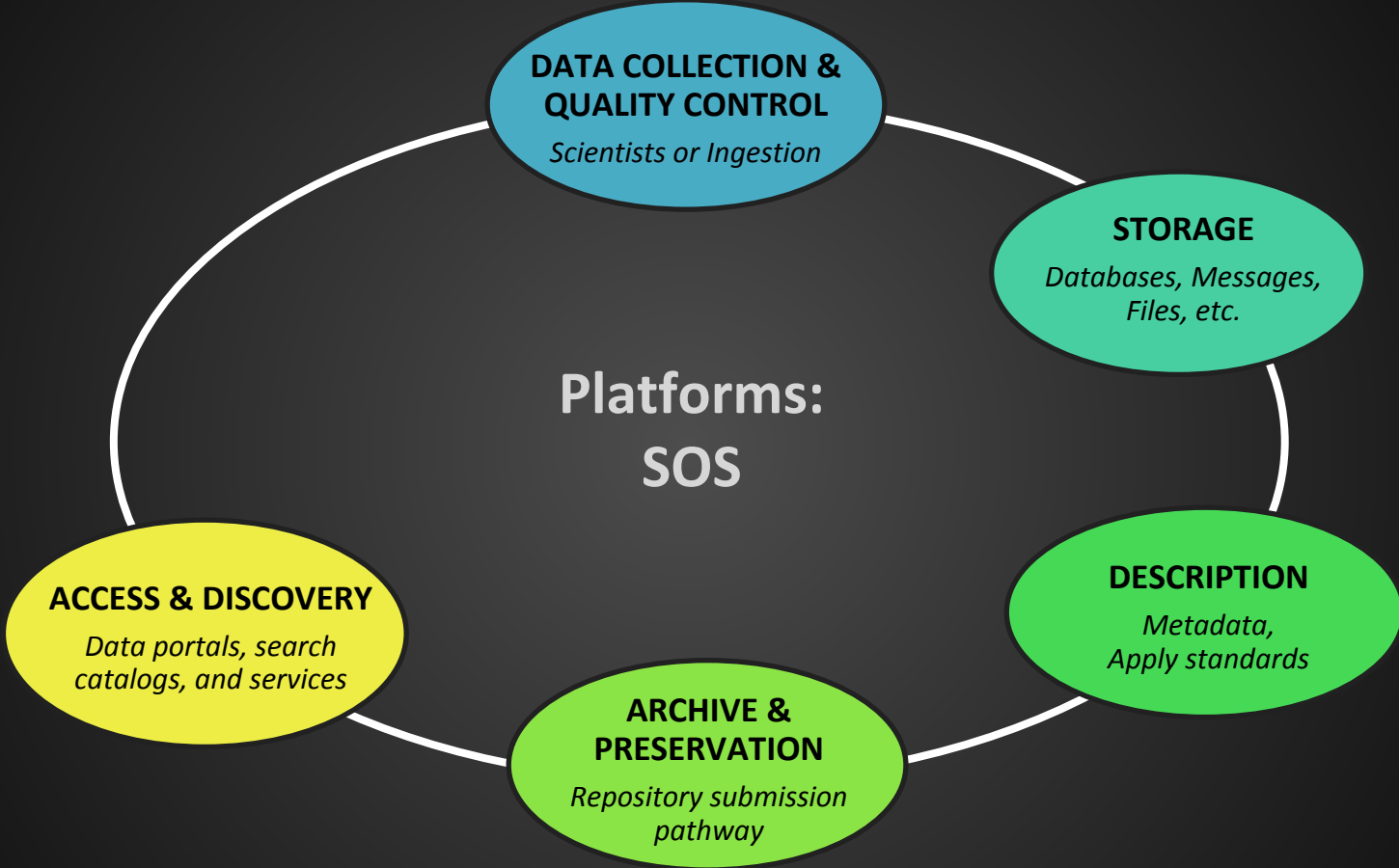
DAC Dataset Flow



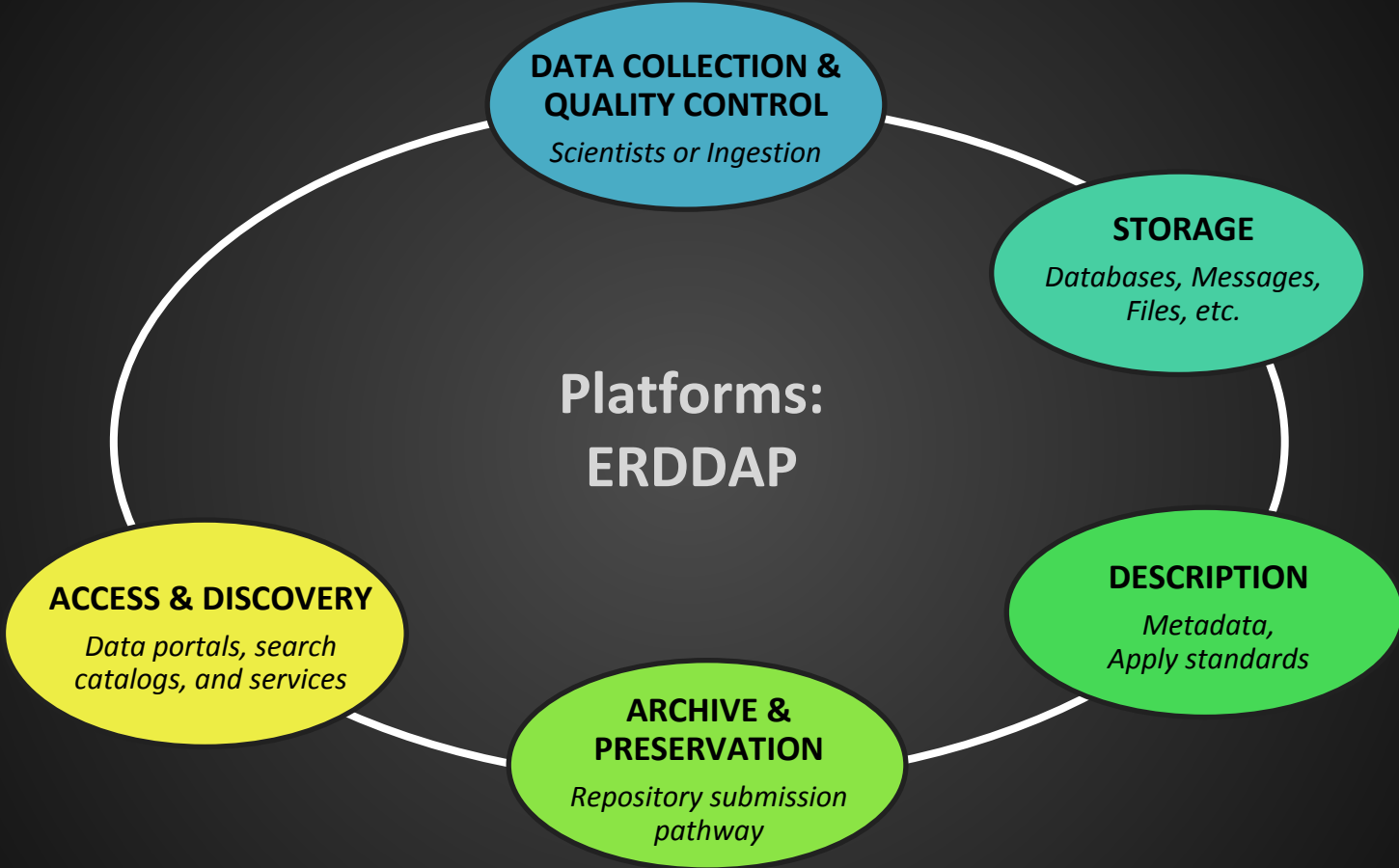
DAC Dataset Flow



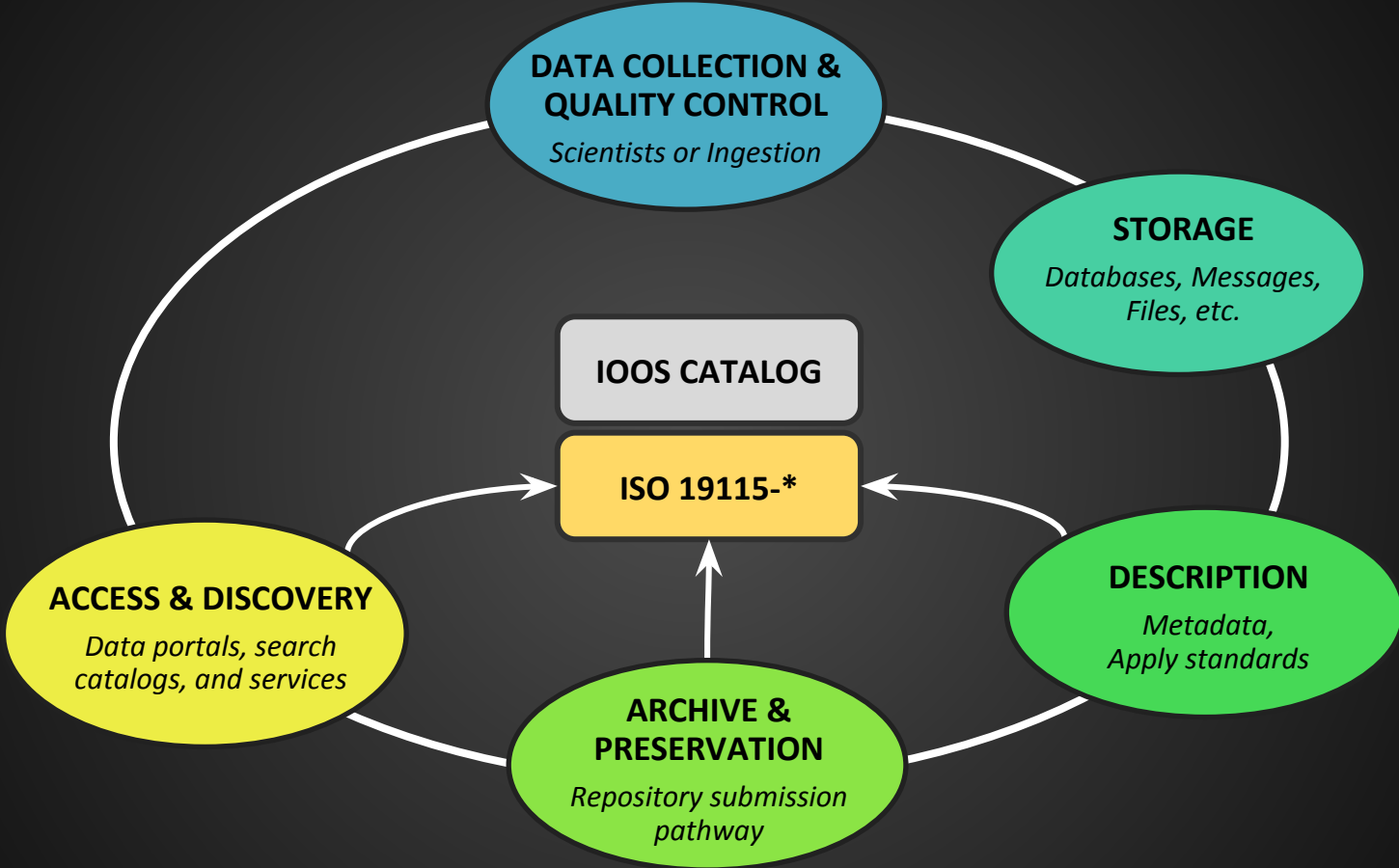
DAC Dataset Flow



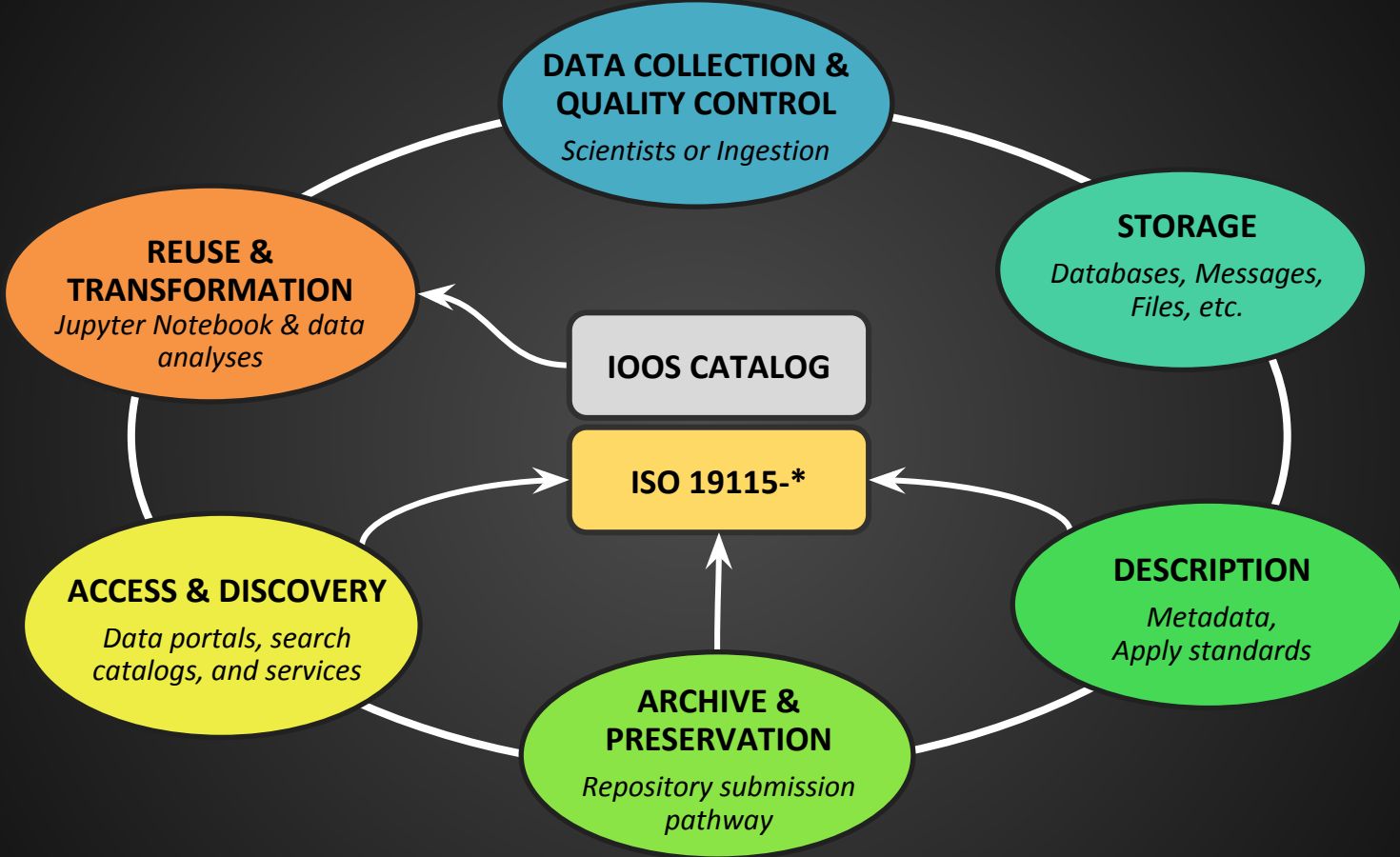
DAC Dataset Flow



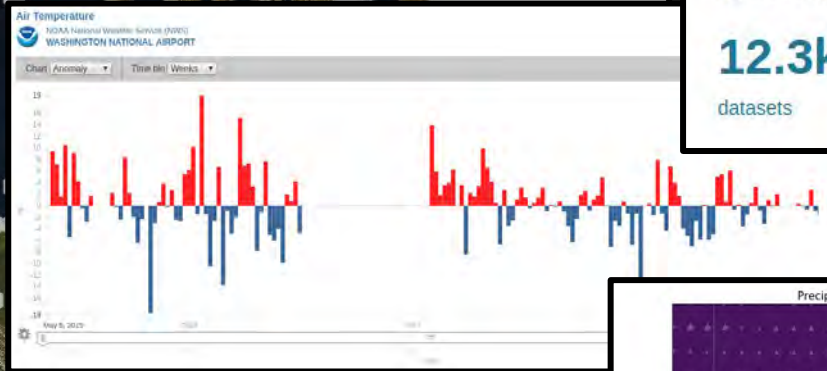
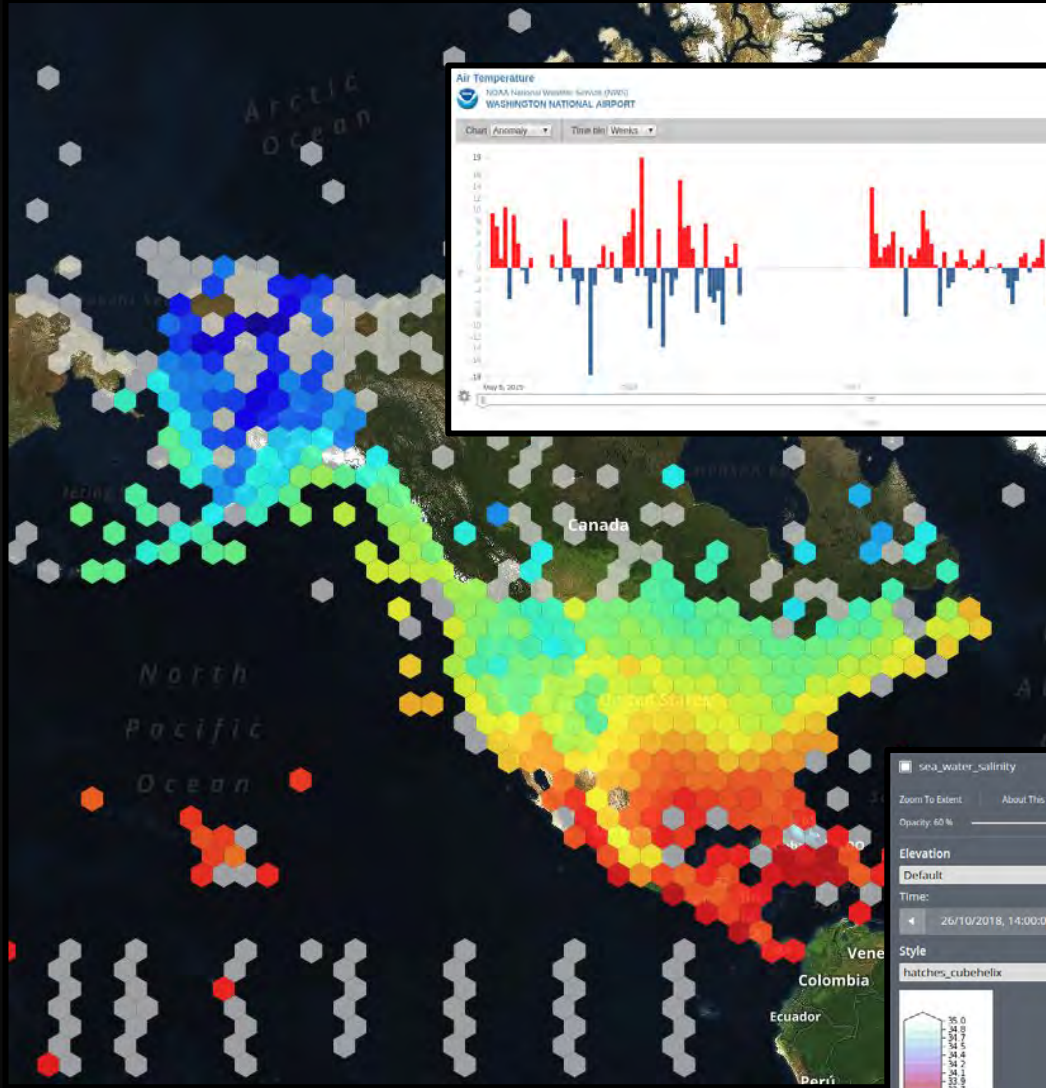
DAC Dataset Flow



DAC Dataset Flow

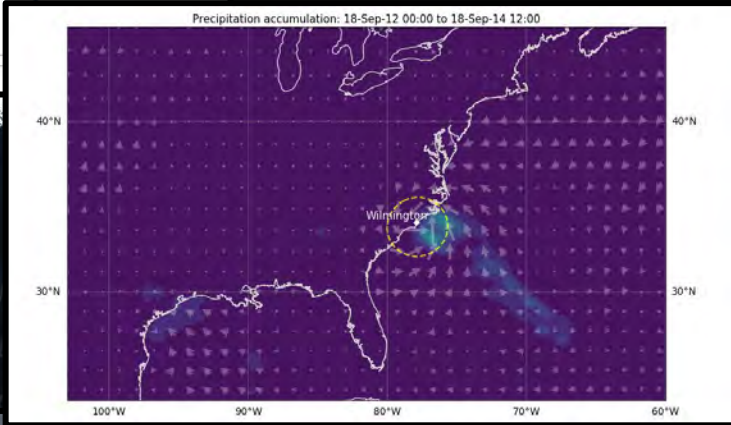


Results



IOOS Catalog statistics

12.3k **22** **0**
datasets organizations groups



sea_water_salinity

Zoom To Extent About This Data Remove

Opacity: 60%

Elevation: Default

Time: 26/10/2018, 14:00:00

Style: hatches_cubehelix



Partners



Exxon Valdez Oil Spill Trustee Council



Questions?

Kyle Wilcox

kyle@axds.co

<https://axiomdatascience.com/>

- IOOS Catalog - <https://catalog.ioos.us>
- IOOS Sensor Map - <https://sensors.ioos.us>
- Northern Gulf of Alaska NSF LTER Site - <https://nga.lternet.edu/>

Extra Slides

HIGH PERFORMANCE COMPUTING (HPC)



- Ocean data is BIG DATA (models, sensors, hydroacoustics, satellite imagery, etc...)
- Requires scalable compute and storage infrastructure

- ~5,000 processor cores
- ~1.5 petabytes of functional storage/5 petabytes of actual storage (~1,500 hard drives)
- Level 2 Fat Tree Infiniband Network, 40 Gb/Sec node to node). 240 Gb/Sec cluster to cluster

