Unleashing the Power of Open Data

NOAA's AI Readiness Journey

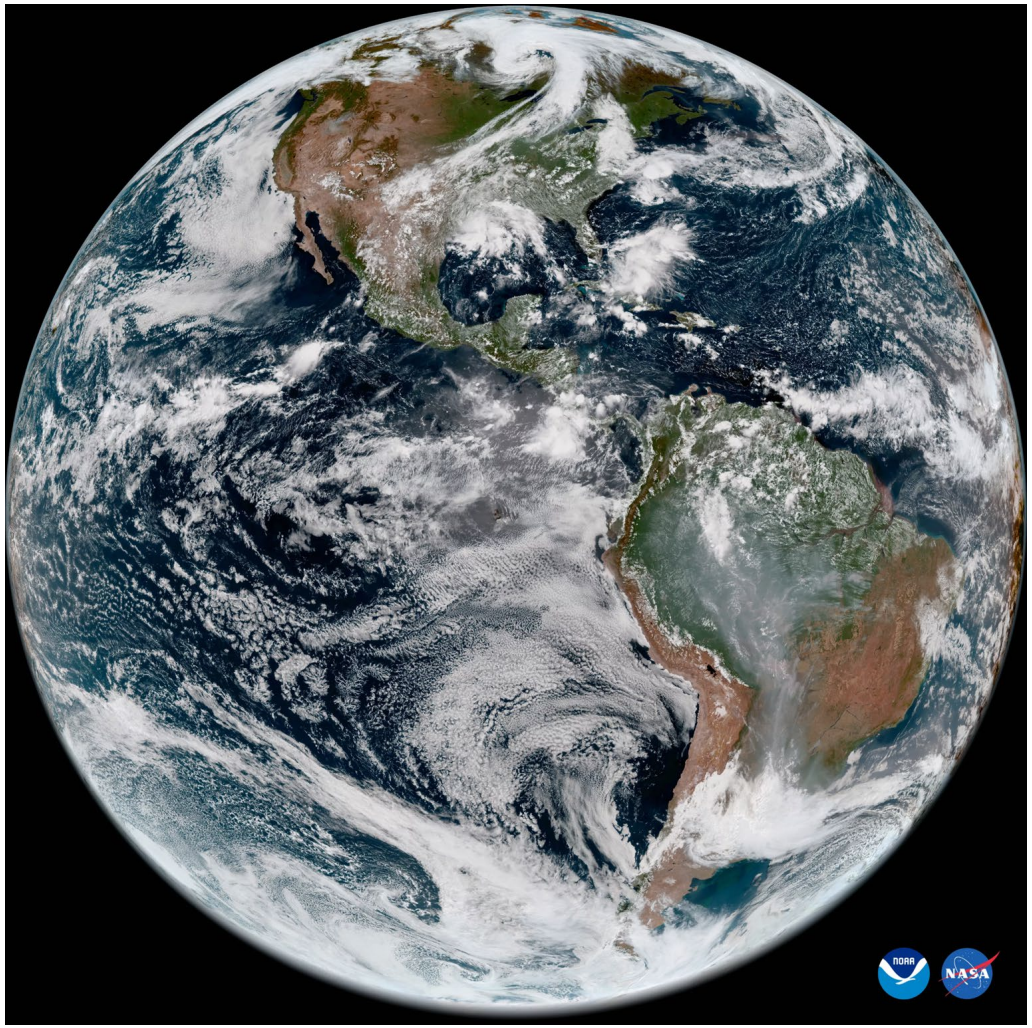
Tyler Christensen, OCS Data Architect

National Environmental Satellite, Data, and Information Service

Sept 2024

Who is NOAA?

National Oceanic and Atmospheric Administration (Department of Commerce)



Our mission

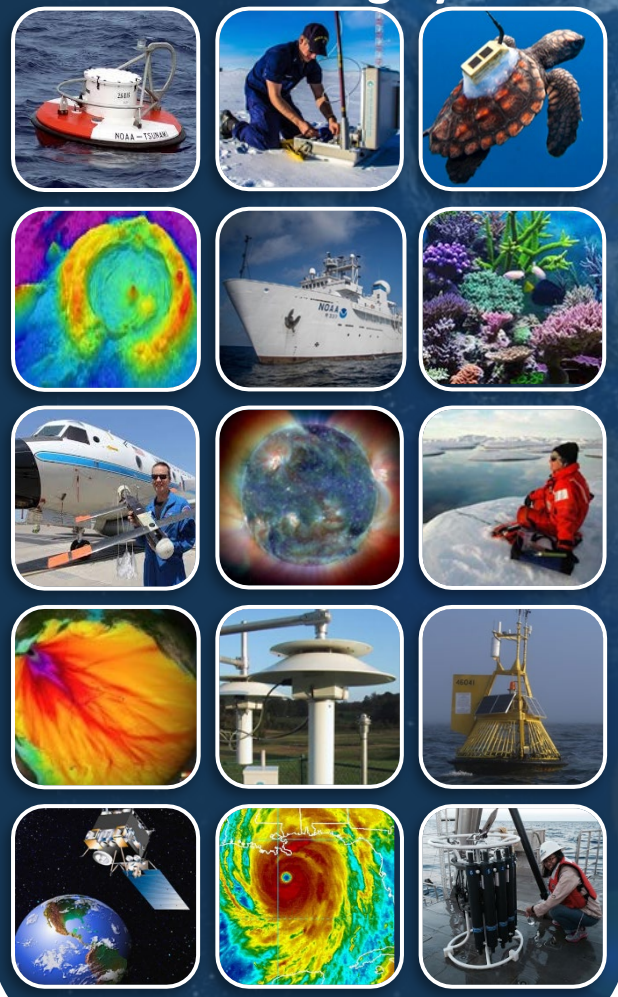
To understand and predict changes in climate, weather, ocean, and coasts,

to share that knowledge and information with others,

and to conserve and manage coastal and marine ecosystems and resources.

NOAA Data Enables You to Make Informed Decisions

Earth Observing Systems

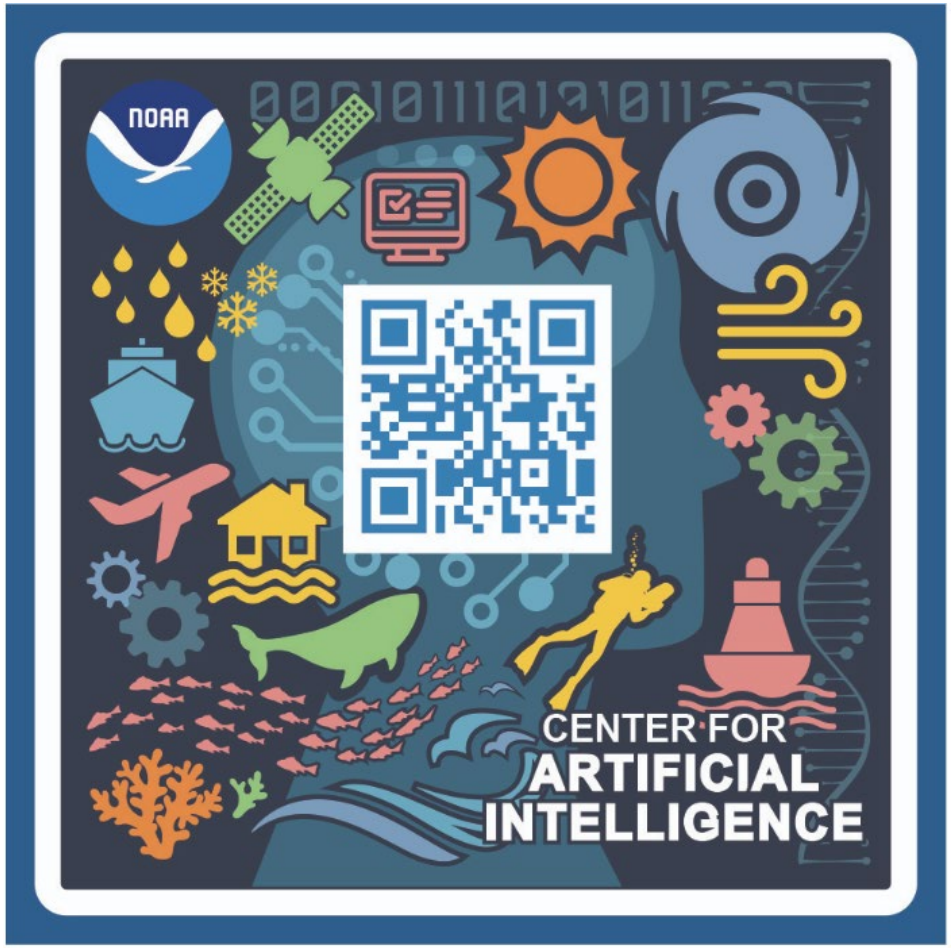


Scientific Data Stewardship

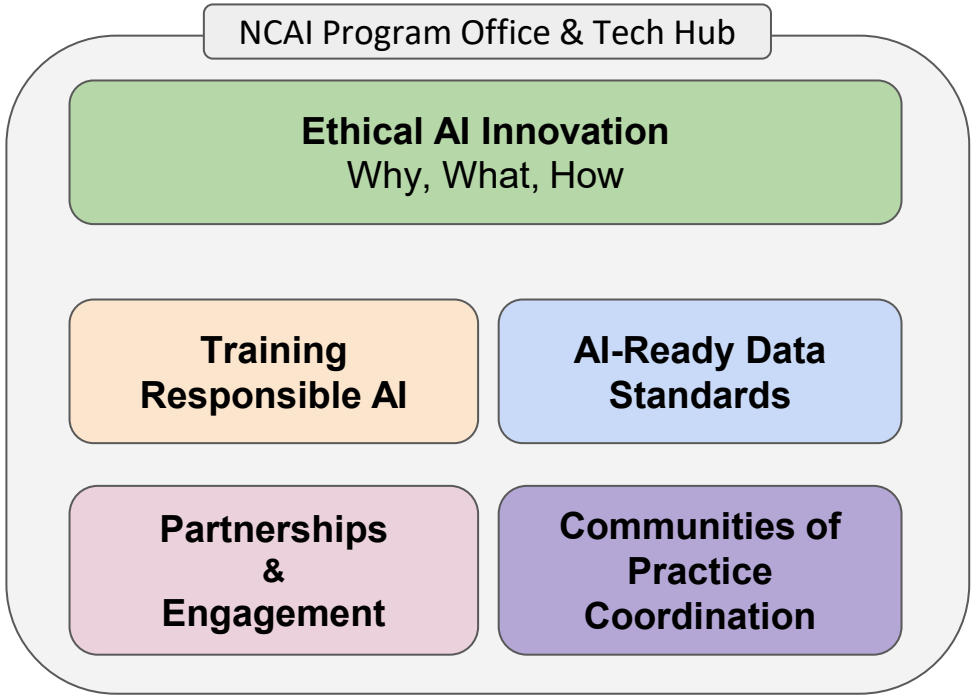
Authoritative Products for Decision-Making

- Snowfall Impact Index — *FEMA*
- Tsunami Warning — *Emergency Managers*
- Heating & Cooling Degree Days — *Energy Sector*
- Hurricane Tracks — *Emergency Planners*
- Coastal Digital Elevation Models — *Hazard Mitigation*
- Solar Activity/Sun Spots — *Power Distribution*
- Annual State of the Climate Reports — *Decision Makers*
- Global & U.S. Climate Summaries — *Numerous Sectors*
- Temperature & Precipitation Outlooks — *Agriculture*
- IPCC & National Climate Assessments — *Gov't Policymakers*
- Billion $ Disasters, Climate Extremes Index — *Insurance*
- Climate Normals — *Construction, Infrastructure, Agriculture*

NOAA Center for AI



noaa.gov/ai



Leading development of a community standard to define AI-Ready Data

https://github.com/ESIPFed/data-readiness

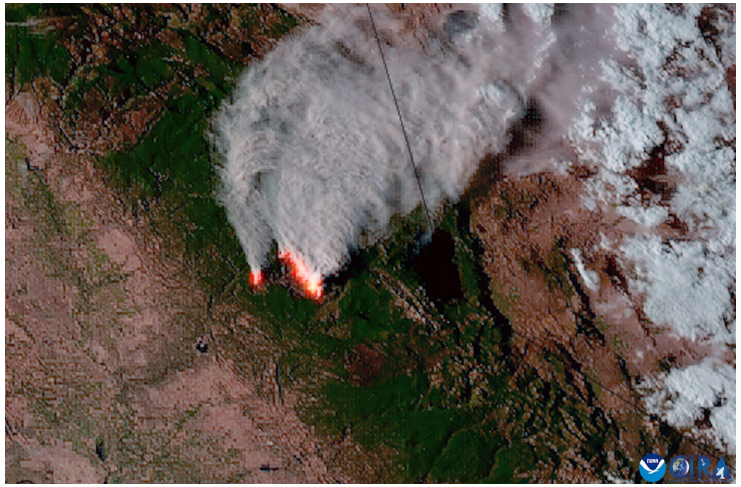
Example AI Applications at NOAA

Developing a drone-based system to detect marine debris and expedite clean-up (link).

VIDEO DATA

Automating the ID of whale calls, speeding up marine mammal research and conservation (link).

ACOUSTIC DATA



Improved detection of lightning from satellites, to enhance fire weather forecasting (link).

GEOSPATIAL DATA

NOAA Has Current Satellite Missions

SWFO
SWFO-L1 - Launches fiscal year 2025

DSCOVR
Operational July 27, 2016

COSMIC-2
Operational Feb. 25, 2020

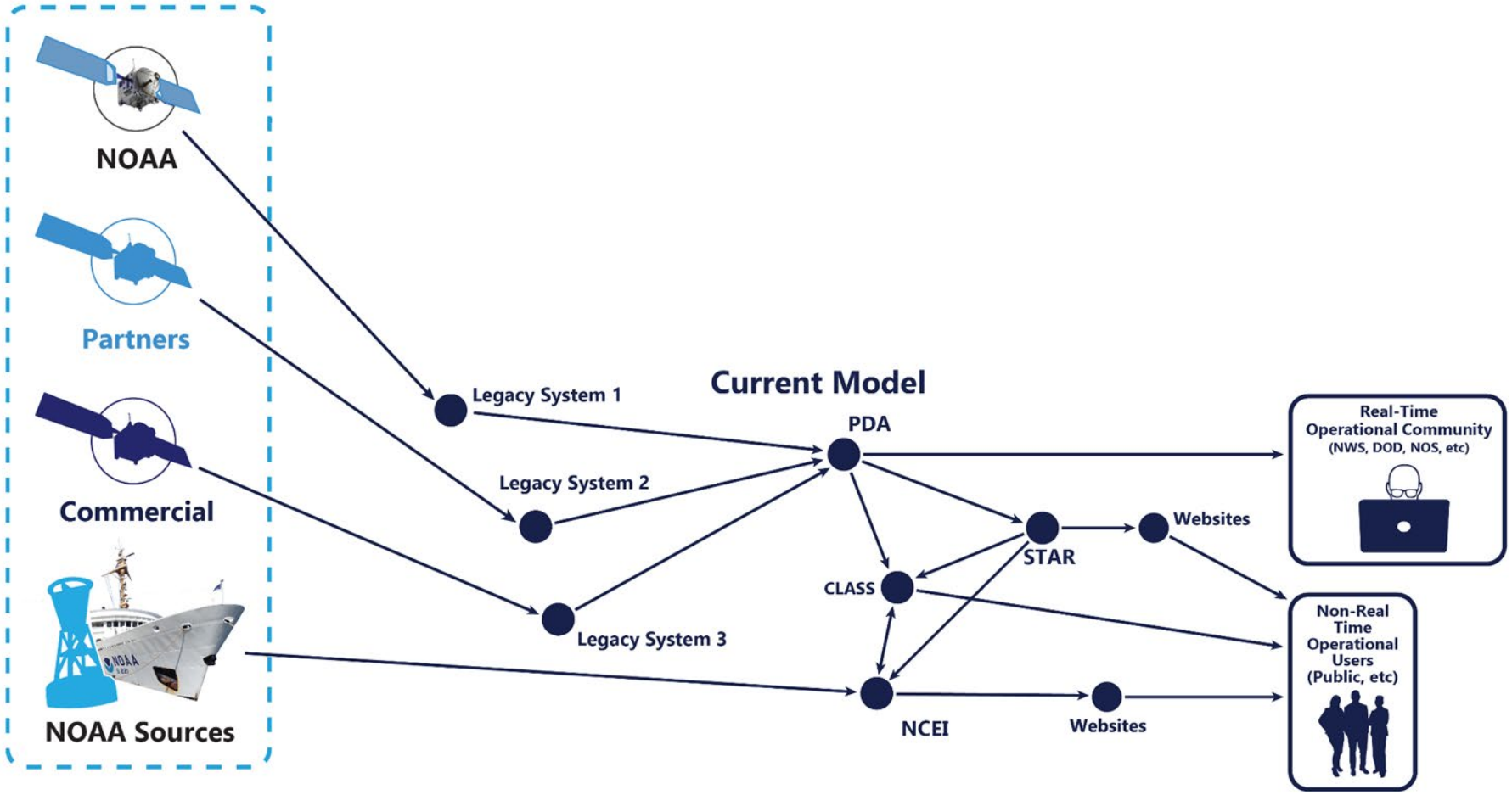
Jason-3
Operational July 1, 2016

GOES-R Series
GOES-16 - Operational Dec. 18, 2017
GOES-17 - Operational Feb. 12, 2019
GOES-18 - Operational Jan. 4, 2023
GOES-19 - Launched June 2024

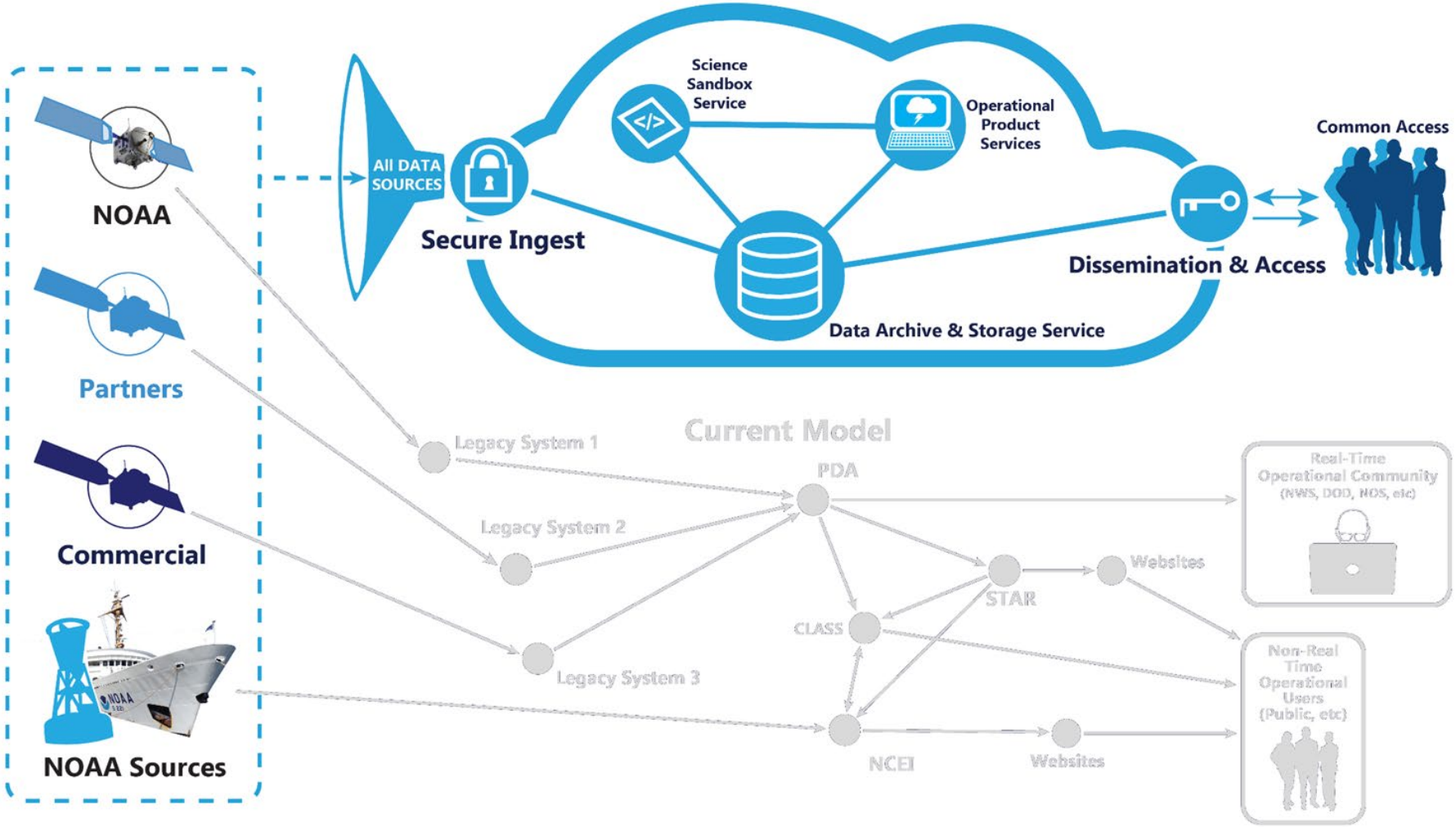
Sentinel-6 Michael Freilich
Operational Nov. 22, 2021

JPSS Series
Suomi-NPP - Operational May 1, 2014
NOAA-20 - Operational May 30, 2018
NOAA-21 - Operational Nov. 8, 2023
JPSS-3 - Launches fiscal year 2033
JPSS-4 - Launches fiscal year 2028

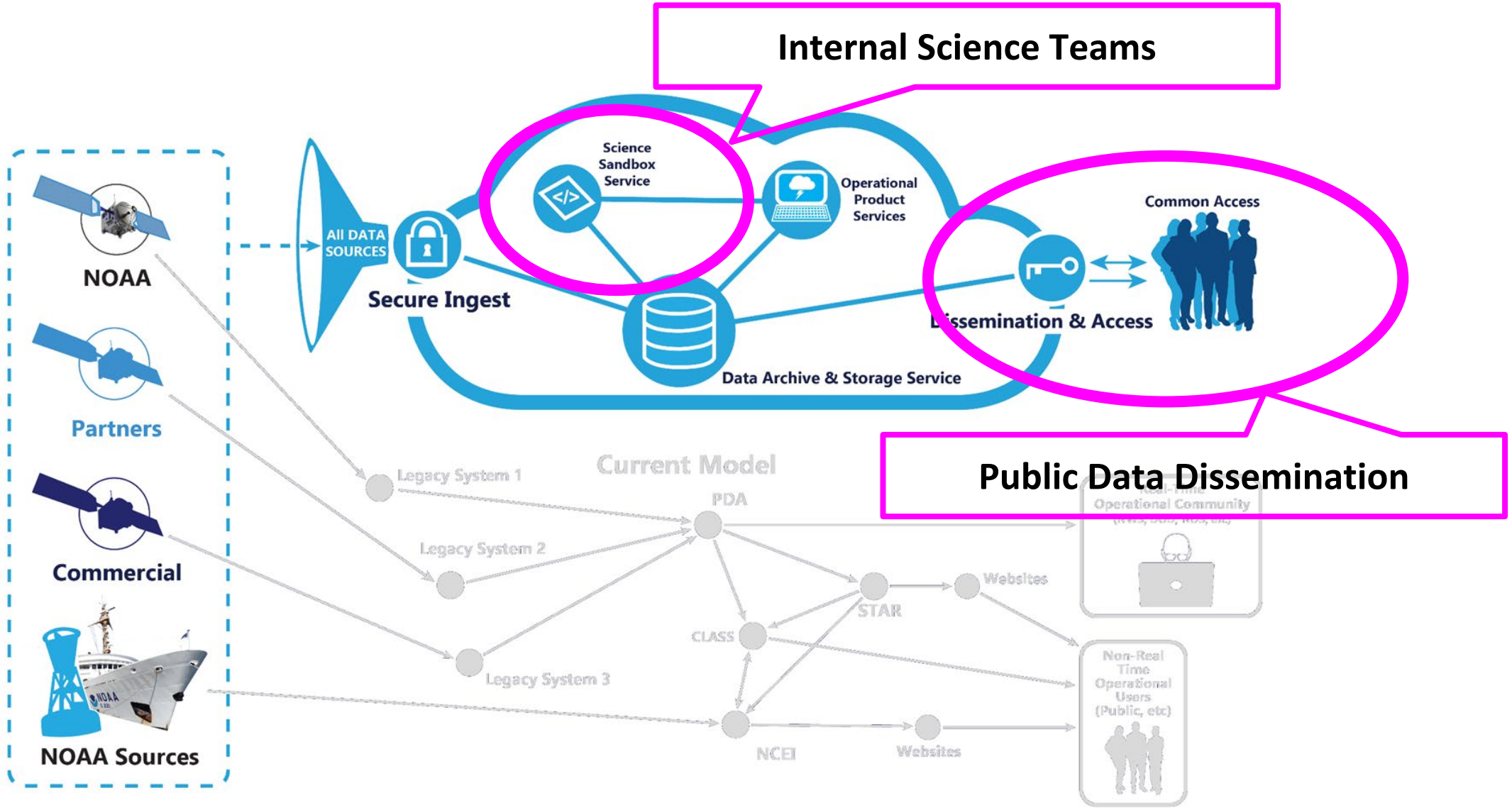
NESDIS Common Cloud Framework



NESDIS Common Cloud Framework



NESDIS Common Cloud Framework



Data Transformation Makes NESDIS' Data More Usable

Enabled by Stakeholder Engagement

01 Develop a cloud-based platform (NCCF) to archive all data and enable broader access based on consolidation

02 Migrate data into the NCCF and transform data to cloud-optimized data formats to enable enhanced access and usability

03 Build an authoritative data transformation toolkit and make it accessible to the public to enable product tailoring by all users

04 Define AI-ready data standards and define a pathway for implementation based on user community needs

05 Transform data to make it AI-ready and aligned with science project implementation plans

Constraints and Opportunities

What are we concerned about?

- Very large data holdings: ~30 Petabytes and growing
  - We don't have unlimited budget for cloud storage & egress
  - NODD, potentially time-constrained
- Different use cases have very different needs
  - genAI needs are different from traditional AI/ML
  - where can we get the biggest return on investment?

What are we excited about?

- Federal agencies have unique data holdings, probably our biggest contribution to AI development worldwide
- Migration to common cloud services is an inflection point

Next Steps

- Developing labeled public datasets for training & benchmarking
  - Example: Tropical Cyclone PRecipitation, Infrared, Microwave, and Environmental Dataset (TC-PRIMED)
- Pilot projects
  - underway for certain public datasets
    - file conversion to ARCO formats like ZARR or Cloud-Optimized GeoTIFF
    - metadata conversion to Spatio-Temporal Asset Catalog (STAC)
    - Kerchunk for existing archives of netCDF files
  - example: Arctic Sea Ice Concentration data for the past 48 years
    - Jupyter notebooks using STAC
- Upcoming pilots for internal science teams

Thank You!