

NOAA National Satellite and Information Service

Creating Data Products in the Cloud: Working in the NESDIS Common Cloud Framework

Melissa Zweng, Kelly Neely, Li Bi, Gian Villamil-Otero, Jianbin Yang, Walter Wolf NESDIS Office of Common Services, February 1, 2024 <u>melissa.zweng@noaa.gov</u>

Summary

- The NESDIS Office of Common Services (OCS) is revolutionizing the way NESDIS provides data and services to its customers
- In the NCCF, users can perform research and development close to the data
- Users will be able to subscribe to and download/use specific data and formats they need using Storefront services
- Unleash NOAA's data for AI/ML and Big Data applications!



Key Challenges

Air Quality





Fire Weather

Flooding





Space Weather



Data and information for a changing world

Remote Sensing

NOAA Data Archive (NCEI)



The NESDIS Office of Common Services (OCS) is a user-focused organization enabling a fully integrated digital understanding of our Earth environment to users



Looking Towards Open Science at NOAA



environmental data



? Open Reproducible Analysis Open scientific

process and reusable analytics pipelines

Open access to research products, results, and new data for reproducibility



NOAA National Environmental Satellite, Data, and Information Service

A Data-Centric Enterprise Cloud Vision for NESDIS

OCS is developing the NESDIS Common Cloud Framework (NCCF), which provides a new concept of operations - to consolidate its operational workloads around centralized data, enable innovative science, and enhance data discovery

- Faster research to operations transition for code updates
- Consolidate and improve access to archived NOAA data holdings
- Innovative science potential with AI/ML tools and centralized data
- Improved system performance through automation and cloud redundancy
- Increased traceability of cost and data use





Data-Proximate Compute: "Science Sandbox" and Ops



NOAA National Environmental Satellite, Data, and Information Service

Data and Product Delivery



Current:

- Operational/real-time PDA on prem
- Delayed mode/research users CLASS, NCEI, STAR

Future:

- Storefront discovery and tailoring services for both public access and custom subscriptions
- Lake House architecture Big Data uses





NOAA National Environmental Satellite, Data, and Information Service

Data Transformation: Cloud Optimized Geospatial Data





NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Al-Ready Data Roadmap



01	Develop and maintain Al-ready data standards	 Short term: Review of common domain data standard Mid term: Develop proposed AI-ready data standard for publishing process Long term: Publish and maintain AI-ready data standard
02	Develop automatic tools for Al-readiness assessment	 Short term: Update assessment form based on user feedback Mid term: Develop metrics to display & synthesize Alreadiness assessment results Long term: Develop automatic Al-readiness assessment tool
03	Develop and improve Al-ready open environmental data	 Short term: Uplift a pilot set of thematic AI-ready data Mid term: Develop tools and leading practices to improve data readiness at scale for community adoption Long term: Provide AI-ready data discovery tools and services
04	Sustain the engagement with user and capacity building	 Short term: Increase the engagement with private sectors Mid term: Develop primers for AI-ready data checklist / standards for different user personas Long term: Develop and maintain training materials on AI-ready data and tools for different user personas

Open collaboration via Github for AI-ready data checklist - <u>github.com/ESIPFed/data-readiness</u>

Next Steps

- 2024 Milestones:
 - Operationalize NCCF Archive Service and the initial Data Access Service
 - Begin migration of archive data holdings into NCCF
 - Baseline NESDIS Data Transformation Strategy
 - Define Cloud-Optimized data standard for NCCF implementation and pilot data transformations
 - Expand AI data readiness in coordination with NOAA Center for AI (NCAI)
- 2025 Milestones:
 - Continue migration of archive data holdings into NCCF, transforming to meet cloud-optimized standards if possible
 - Expand AI data readiness in coordination with NCAI
 - Potential to fund an AI science project that leverages existing AI-ready data in NCCF



Takeaways

- The NESDIS Common Cloud Framework (NCCF) is revolutionizing the way NESDIS provides **data and services** to its customers
- In the NCCF, users can perform research and development close to the data
- Users will be able to subscribe to and download/use specific data and formats they need using Storefront services
- NOAA seeks to:
 - Grow our Data Communities of Practice by developing best practices and learning from others
 - Build AI-ready datasets to meet user needs
 - Be a hub for innovation; engage and support **pilots and research**

