PREPARING FOR A CLOUDY FUTURE

Progress Update on Report Recommendations Provided by the Data Archive and Access Requirements Working Group (DAARWG)

Submitted to NOAA through the Science Advisory Board (SAB) – December 2019

March 15, 2021

Tony LaVoi, Acting NOAA Chief Data Officer

Purpose

- Provide a progress update on NOAA's implementation actions in response to the SAB's Cloudy Future report recommendations

Topics

- Current context: state of Cloud adoption and Science and Technology (S&T) planning in NOAA

- Cloudy Future recommendations alignment with NOAA S&T and Information Resources Management (IRM) Strategies

- Selected examples of progress

State of NOAA Cloud Adoption

- **Planning**
  - NOAA Cloud Strategy and Implementation Plan
  - NOAA S&T Strategies (5) - all leverage NOAA cloud and data enterprise
  - Information Resource Management (IRM) Strategy & Implementation Plan
- **Execution**
  - NOAA Cloud Committee and Working Groups
  - Contracting (Cloud Utility Contract, Big Data Program, High Performance Computing)
  - Enterprise Software-as-a-Service IT System
  - Security: Common Security Controls; Cloud Access Security Broker
  - Cloud adoption in all NOAA line offices
- **Innovation**
  - NESDIS Common Cloud Framework
  - Google Artificial Intelligence (AI) Testbed
- **Training and Support**
  - Cloud Service Provider Summits
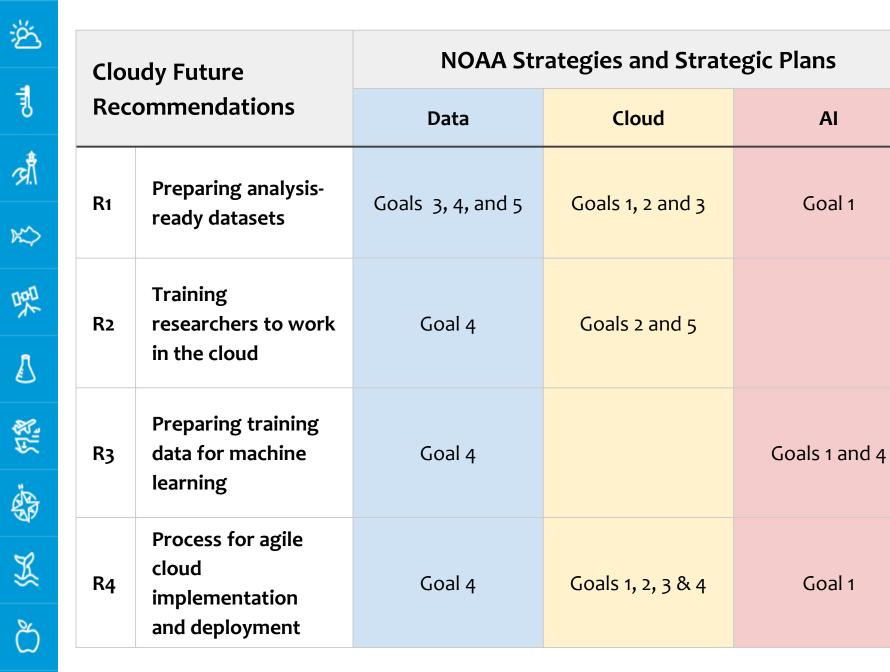 - Workforce training and development

NOAA S&T Strategies

Cloudy Future Recommendations	NOAA Strategies and Strategic Plans		
	Data	**Cloud**	**AI**
R1 **Preparing analysis-ready datasets**	Goals 3, 4, and 5	Goals 1, 2 and 3	Goal 1
R2 **Training researchers to work in the cloud**	Goal 4	Goals 2 and 5	
R3 **Preparing training data for machine learning**	Goal 4		Goals 1 and 4
R4 **Process for agile cloud implementation and deployment**	Goal 4	Goals 1, 2, 3 & 4	Goal 1

Cloudy Future Recommendations		NOAA Examples
R1	Preparing analysis-ready datasets	• AI-Ready Data Standard (R1.1) • NOAA Big Data Program (R1.1) • Google OTA (R1 & R3)
R3	Preparing training data for machine learning	
R2	Training researchers to work in the cloud	• AI Proficiency in the Workforce (R2) • HPC Hackathon (R2.4)
R4	Process for agile cloud implementation and deployment	• *Considerable amount of Agile software development approaches being used across the organization*

AI-Ready Data Standard

- Federal agencies need to improve the AI-readiness of their open data -- including training datasets ([White House 2019 memo](#))
- No standards exist to define the requirements, just [an unpublished first draft](#) from an Office of Science and Technology Policy subcommittee
- NOAA strategies for AI, Cloud, and Data require us to define the criteria

- NOAA is building a government / academic / NGO / industry collaboration team [via the Earth Sciences Information Partnership (ESIP)](#) (March) -- includes access to Amazon Web Services (AWS) credits
- Aiming for 80% completion of the standard in 2021
- Long-term goal to establish a formal standard via NIST / ISO / etc.

Big Data Program Current State

- Transitioned from Cooperative Research and Development Agreements (CRADA) to contracts with Amazon, Google, and Microsoft; operational NOAA program

- Stability provided through a 10 year contract (5 x 2 years/period)

- Approximately 8 PB of NOAA data publicly accessible through 3 Cloud Service Providers - over 145+ datasets in the cloud

- Limited data transformations to date; strong desire to identify resources to support cloud-native formats

- BDP has enabled a significant increase in data usage, supporting users and decision makers across various sectors of the economy and research community

Google AI /ML Agreement

What is it?: In September 2020 NOAA and Google entered into an Other Transactional Authority (OTA) agreement that leverages NESDIS science and IT/cloud subject matter experts with Google's AI/ML world leading practitioners to explore if these tools can enhance weather forecasting and research capabilities

One: Explore if a relationship like this can yield benefits for NOAA

Two: Determine if we can enhance our weather forecasting and research capabilities with AI/ML tools & techniques; potential benefits:

- Gain direct exposure and training from Google AI/ML SMEs that may enhance our research capabilities

- Jointly develop usable code – NOAA will have an unlimited use of that source code

- Gain insights into how to better address our challenges with Big Data in 'nowcasting' and 'numerical weather prediction'

Building AI/ML/Cloud Proficiency in the Workforce

- All S&T strategies and IRM strategic plan demand a diverse and inclusive workforce to reflect, understand, and respond to the varied communities and stakeholders NOAA serves
- Workforce development is a common goal across all strategies, examples include
  - **IT Workforce Strategic Plan**
  - Planning for the **"Future of Work"**
  - NOAA Center for Artificial Intelligence (AI) will serve as the **NOAA Learning Center** to provide critical AI training resources and expand training opportunities for employees across NOAA (upskilling of individuals, external training opportunities)
  - NOAA Office of Education exploring the creation of **student internships, training, and experiential research opportunities** for undergraduate and graduate students, including students from diverse backgrounds and underrepresented groups
  - Leverage **joint expertise, optimize collaborative investments,** and facilitate scientific and technical information exchange between NOAA LOs and other organizations

Hackathons

- NOAA together with NVIDIA hosted the first virtual Graphics Processing Unit (GPU) Artificial Intelligence (AI) Hackathon on December 7-9th

  - GPU Hackathons provide exciting opportunities for scientists to accelerate their AI research or HPC codes under the guidance of expert mentors from National Labs, Universities, and Industry leaders in a collaborative environment

  - NOAA AI Hackathon was a multi-day event designed to help teams of three to six developers accelerate their own code on GPUs using a programming model, or machine learning framework of their choice

- The NOAA Citizen Science strategy cites challenge competitions and Hackweek style events as a best practice and an important practice to address infrastructure issues and science challenges

Summary

- Provided a progress update on NOAA's implementation actions in response to the SAB's Cloudy Future report recommendations

  - *Current context: state of Cloud adoption and S&T planning in NOAA*

  - *Cloudy Future recommendations alignment with NOAA S&T and IRM Strategies*

  - *Highlighted examples of progress*

AI Highlights

- **AI-Based Enterprise Algorithms for Satellite Data Exploitation**
  - NESDIS Center for Satellite Applications and Research has been developing efficient (AI-based) enterprise algorithms for a variety of products generated from space-based observations taken from both polar and geostationary satellites

- **CoralNet: Automated, Human-in-the-loop, Point Annotation of Coral Reef Imagery**
  - CoralNet is an operational AI enabled application to efficiently annotate coral reef images in support of the NOAA National Coral Reef Monitoring Program mission. NOAA Ocean and Fisheries Services have worked collaboratively with its academic partners to improve the CoralNet architecture with new ML capabilities and expanded training data, decreasing the error rates by 22% over the operational product

- **High-Resolution Land Cover Mapping**
  - NOAA's National Ocean Service Coastal Change Analysis Program (C-CAP) has used traditional machine learning (ML) methods since 2005, and recent advances in deep learning approaches have improved imagery resolution (from 30-meter to 1-meter) enabling the delivery of higher quality products nationally. NOAA uses ML convolutional neural networks (ConvNets) for mapping coastal, agricultural and other land cover features. NOAA continues to work with AI experts in the private sector to cost share large-scale continental U.S. (CONUS) impervious surface mapping

- **Monitoring of Ocean State for Improved Climate Predictions**
  - OAR and CIMES (Princeton) scientists are employing an AI/ML framework to recognize key ocean dynamical regimes.

NESDIS Common Cloud Framework (NCCF)

The NCCF is a suite of services that allows NESDIS to provide end-to-end ground capabilities
- Cloud agnostic common enterprise architecture that enables NESDIS business functions
- Provides secure infrastructure, tools, and services to meet NESDIS user needs