

DAARWG recommendations concerning NOAA Data and Cloud Plans

Prepared by

Data Archive and Access Requirements Working Group
of the NOAA Science Advisory Board

Co-chair: Dr. J-F de La Beaujardière, NCAR

Co-chair: Dr. Molly Jahn, DARPA

Mr. Thomas Huang, NASA/JPL

Dr. Ana Pinheiro Privette, Amazon

Dr. Joe Hamman, CarbonPlan & NCAR

Dr. Kerstin Lehnert, LDEO

Mr. Shane Glass, Google

Col. Mike Gremillion, U. Alabama Global Water Security Center

Dr. Lucas Joppa, Microsoft

Summary

- DAARWG thanks NOAA for opportunity to comment on two plans:
 - "NOAA Data Strategy Implementation Plan - Draft Narrative and Actions" (Jan 2021)
 - "NOAA Cloud Strategic Plan - Draft V3" (Jan 2021)
- Focused on the "actions" rather than on previously-approved Goals and Objectives
- **Pleased to see NOAA continuing to give Data and Cloud matters serious consideration**
 - WG does have some recommendations for possible improvement
 - Will discuss Data plan first, then Cloud plan

Data Plan – General Comments

- Goals and Objectives worthwhile and supported by the Actions
- Applaud continued focus on open data sharing and use of data standards
- Pleased to see an explicit call for stakeholder engagement
- Planned actions are achievable if NOAA provides the resources to complete them
- Note large number of Actions (22) and associated Tasks (65) in the Plan.
 - No target due dates or responsible parties/groups listed

Data Plan – Recommendations on Action 1.1

- *Action 1.1 Establish the NOAA CDO organizational structure.*
- **Recommendation 1:** The Plan should state who ultimately owns and evolves the Data Strategy for NOAA -- is it the CDO, or another individual or body?
- **Recommendation 2:** The Plan should clarify the source of the CDO's budget and how it can be protected from cuts in future years.

Data Plan – Recommendations on Action 1.2

- *Action 1.2 Ensure the NOAA CDO is represented on the appropriate NOAA decision making bodies [that have "data connections"]*
- Nearly every NOAA council & committee has some sort of "data connection"; don't stretch the CDO too thin
- **Recommendation 3:** Carefully consider which groups should actually have CDO representation, and ensure all groups follow existing and future data-related policies.

Data Plan – Recommendations on Action 2.1

- *Action 2.1 Establish a NOAA Data Governance Body*
- Having both Data Governance Body (DGB) and Environmental Data Management Committee (EDMC) may be confusing and counterproductive
- **Recommendation 4:** Either (a) very clearly delineate the scope and relationship of DGB and EDMC, or (b) rename EDMC as DGB, or (c) recharter EDMC to include the functions and authority intended for DGB.
- **Recommendation 5:** If a separate (non-EDMC) body is to be created, consider following established NOAA nomenclature and having it be called the Data Council; EDMC could then report to the Data Council instead of to both the NOAA Observing Systems Council and CIO Council as currently done.
- **Recommendation 6:** Consider whether the CDO should be Chair *ex officio* of the Data Council, DGB, or EDMC, as appropriate.

Data Plan – Recommendations on Action 3.2 & 3.3

- *Action 3.2: Publish and promote a NOAA Open Data Plan.*
- *Action 3.3: Maintain a comprehensive NOAA data inventory, including standard metadata, that is published to NOAA and Federal data catalogs (data.gov)*

- NOAA Plan for Public Access to Research Results (PARR Plan) already issued in 2015
- NOAA Data Catalog already established in 2016 (<https://data.noaa.gov/dataset/>) and harvested by data.gov

- **Recommendation 7:** The Plan should avoid duplicating or rewriting past policies unless there is an identified shortcoming with the prior documents.
- **Recommendation 8:** The Plan should avoid duplicating past actions or reinventing past methods unless there is an identified shortcoming with the prior approach.

Cloud Plan – General Comments

- DAARWG applauds the desire to increase use of the Cloud by NOAA.
- The Goals and Objectives are generally good.
- The Actions have specific target dates and responsible parties, which is a plus.

- Some of the Actions in the Plan are fundamental and should be focused on, notably:
 - Action 1.1.4: Establish an **exploratory environment**, with guardrails, for cloud service tests and trials.
 - Action 1.2.1: Establish a **Community of Interest**, composed of skilled cloud project managers.
 - Action 1.3.2: Create an environment where "**smart failures**" are perceived as a positive outcome
 - Action 2.4.3: Create a **pilot project** for a cloud development environment that moves both the processing and developing test data to the same cloud environment.

Cloud Plan – Agility recommendations

- Very large number of actions - 37 in FY2021, 31 in FY2022.
A prescriptive plan with many steps is the opposite of "agile."
- Plan seems overly focused on cost and security rather than actually deriving benefit from the cloud; bureaucratic and cautious rather than agile.
- **Recommendation 1:** Highlight and prioritize key Actions relative to other actions
- **Recommendation 2:** Consider how to quickly establish an exploratory environment (Action 1.1.4) as soon as possible.
- **Recommendation 3:** Reduce or consolidate the number of actions
- **Recommendation 4:** Indicate how the plan can be adjusted during execution for greater agility

Cloud Plan – Training Recommendation

- *Action 5.1.1: "... provide access to cloud training resources by the CSPs [Cloud Service Providers] for project managers"*
- Rank-and-file software engineers also need training, and should be permitted to spend work hours and project funding on training
 - See recommendation #2 of *Cloudy Future* report
- **Recommendation 5:** Provide a greater emphasis on training for diverse constituents.

Cloud Plan – Benefits & Metrics Recommendations

- The Plan does not clearly identify some of the potential benefits of using the cloud for NOAA and users of NOAA data, such as being able to compute directly on high-volume data stored in the cloud without moving the data, or using vendor-provided managed services to reduce cost and implementation effort.
- Other federal agencies also have data of relevance (not just "lessons learned")
- Not clear how the impact and success of this Plan will be measured.

- **Recommendation 6:** Include discussion of the how NOAA might leverage the more unique benefits of the Cloud beyond mere infrastructure.
- **Recommendation 7:** Indicate whether collaboration on data co-location with other agencies would be desirable and possible.
- **Recommendation 8:** Indicate what constitutes progress or success for NOAA in its Cloud adoption.

Questions? Discussion?

J-F de La Beaujardiere, PhD - DAARWG Co-Chair

National Center for Atmospheric Research

jeffdlb@ucar.edu