

# EPA Data Governance Council

Meeting Summary – July 31, 2023

## General CDO Updates

- Jan Krysa has taken a new opportunity within Office of Research and Development and will be stepping down from Chair of the AI Subcommittee.
- The memo announcing the Data Officer program has been distributed. Data Officer assignments should be completed by August 25. The Data Officer program will be holding the first meetings after Labor Day.
  - The first focus for the Data Officer group will be to set priorities for Data Governance tasks.
  - The second focus for the Data Officer group will be to identify individuals performing Data Stewardship roles related to the Data Lifecycle.
- There are several Office of Management and Budget and Office of Government Administration groups working to establish guidelines for the federal government on several AI topics including ethics, and security. The governance related to Large Language Models, such as ChatGPT, is still being considered. Those applications will remain off limits for now. By August, EPA may have established more AI governance.
- The Data Communications Subcommittee is currently redesigning [EPA.gov/data](https://www.epa.gov/data) to modernize and upgrade the site with a data-centric focus.
- Progress is being made in the Amazon Web Services Open Government Program. EPA has signed the memorandum of understanding with AWS. A team is working out the process to ensure all the data is transferred legally, efficiently and without losing any data. The current subcommittee is called CRADA but that may be changing. Anyone interested in housing a data set in AWS can work with this group to ensure the data set meets the established criteria.
- The Information Creation Request process is being revised to improve the process and capture how and where data is being collected for all EPA projects.
- A team is working to improve Enterprise Data Management to ensure that system solutions being implemented are necessary and appropriate.

## Status Updates from Subcommittees

- Aaron Zull discussed sunsetting the Data Stewards Network Formation Subcommittee.
  - The approach to define the Data Steward term and role has changed to be based on the Data Life Cycle. The Data Stewards Network Formation Subcommittee will be dissolved.
  - Future work for defining the Data Life Cycle and related Data Steward roles will be completed by the Data Officers group.
  - Data Officers will identify individuals that are already performing the role of Data Steward instead of assigning the Data Steward title in conjunction with new responsibilities.
- Anna Green presented updates from the Regulated Facilities Governance Subcommittee.
  - The first phase of the Regulated Facility Data Needs Project was completed in 2020 and utilized the collection of 200 survey responses and 23 interview groups to identify the top six facility data business needs.

- Phase 2 of the project was completed between 2020 and 2022. During this phase, workgroups were formed with representatives from EPA, States, and Tribes to perform detailed analysis of EPA data processes.
- The project concluded with a recommendations paper. The primary recommendation was the establishment of governance for Regulated Facility Data.
- The Regulated Facilities Governance Subcommittee has been formed to create, implement and maintain the required governance structure.
- Goals, actions, products and expected challenges were identified for the subcommittee and a charter was drafted.
- The next steps to form the subcommittee are to finalize the charter and to recruit members.
- The kick-off meeting for the Regulated Facilities Governance Subcommittee will be at the end of August or beginning of September.

## Data Skills Training and Data Inventory / Alation Updates

- Janet Kremer presented updates from the Data Skills Training and Data Inventory Subcommittees.
- Based on the analysis of the Data Skills Survey in 2021, the Data Skills team has developed Data Personas for Early Learner/Beginner, Novice, Data Savvy, and Advance Data Savvy staff.
- A new Data Skills page for the Early/Learner Persona is available on FedTalent. It will contain existing training resources that apply to this skill level.
- The Data Skills Team will focus on developing training for the Novice Persona next.
- The Data Inventory Subcommittee has been shifting to a more mature model to align data steward roles and responsibilities with data lifecycle stages.
- In the next step, DIS will work with Data Stewards to implement use cases.
- Alation is the new enterprise metadata management system being established within EPA. The Enterprise rollout is scheduled for the first quarter of 2024. All employees will have access, but Data Officers and data stewards will have editing capabilities.
- Alation will not be publicly accessible but EPA data will be available through Data.gov.
- The Environmental Dataset Gateway search interface has been decommissioned. Staff may visit Data.gov to reach EPA records until Alation is available.
- The EPA Open Data Metadata Editor, EPA's Clip & Ship and Science Hub will not change.

## Common Data Language in EPA Contracts

- Kathryn Krueger presented on the common language that should be included in EPA contracts to ensure data collected or used within EPA projects is the property of EPA and is placed in the correct locations in the correct format.
- The language that has been drafted includes the requirement that contractors:
  - Create a data management plan.
  - Define data management responsibilities.
    - Identify specific Unmanned Aerial System data management responsibilities.
    - Identify specific Geospatial data management responsibilities.
- The next steps to improve data collection and standardization may include:
  - The creation of a clause that would be required in any new contract that involves data.

- The creation of a DGC Subcommittee that can more clearly identify collection standards for different kinds of EPA data.
- Collecting the existing language that has been developed across the agency for specific types of data.
- Working with FITARA to add this required data language identification in their review procedure.