

EPA Data Governance Council

Meeting Summary – May 20, 2020

Teams Site Overview

- EPA's Chief Data Officer (CDO) provided an overview of the Data Governance Council Teams site, which the Council will use to support most of its work.
- Council members are encouraged to document their organizations' data challenges. This will help the Council prioritize activities.
- The site will be used to document the Agency's data assets. The Council will be asked to review those assets this summer to ensure accuracy.

Recap of Some Activities to Date

- The CDO recapped the activities to date.
- 2020 Data Action Plan
 - OMB has identified six major actions each agency must at least start, if not complete, in 2020:
 - Identify data needs to answer priority Agency questions. This will be a part of the learning agenda eventually.
 - Institutionalize the Agency's data governance model. There is some data governance in place, including this Council, but it is not uniform across the Agency. EPA will need to have at least a plan for the data governance model by the end of the year.
 - Assess the Agency's data and related infrastructure maturity.
 - Identify opportunities to increase staff data skills.
 - Identify EPA's priority data assets for the Agency's needs and for the open data plans.
 - Publish and update the data inventories.
- January 2020 Information Technology (IT) / Information Management (IM) Conference
 - During the Conference, Agency leaders had conversations about EPA's data challenges and opportunities to better connect data assets.
 - Some of the takeaways from the meeting include:
 - EPA needs to do a better job at exposing and sharing its data across the Agency.
 - EPA needs to improve tagging and metadata for unstructured data (e.g., assets on OneDrive) to improve discoverability.
 - EPA needs to ensure the Agency has staff that know how to manage and use EPA's data properly.
 - EPA needs to enable better service to the data, both internally and externally.
 - At the Conference, the CDO outlined his vision for data's place within EPA: data is the bridge between business needs and IT assets so it deserves a central, equal place at the table.
- IT Portfolio Reviews (ITPRs)
 - The CDO has been meeting with each region and program office to discuss their current data inventory, priorities, and challenges.
 - These conversations will be used to document EPA's data inventory for OMB as well as prioritize activities.
 - OMS has requested is that each agency identify those resources that can be shared across the federal government to address the COVID-19 response.

Walkthrough of Upcoming Activities

- The CDO provided an overview of upcoming activities, many of which are OMB mandates.
- Data Inventory
 - OMB has mandated EPA document its data inventory.
 - EPA's current data inventory is housed in the Environmental Dataset Gateway (EDG), but the data assets are not prioritized. The EDG also needs to be updated or migrated.
 - EPA's goal in 2020 will be to capture as many of the critical data assets that may be missing and add them. The CDO will use information learned from the ITPRs to identify the inventory.
 - EPA's goal is to understand: what the Agency has, how it is resourced, how much storage space is used, and how much money and people hours EPA spends managing them. This information will be critical to help plan in case of any space and funding reductions.
 - The Governance Council should look at what data the Agency is using, how much of the data are copies, and why users are creating copies. This could help the Agency find out ways to better structure the data and make it more usable.
- Priority Data Assets
 - As part of the Data Inventory activities, the Agency will need to identify its priority data assets.
 - This will include defining a process to determine and assign authoritative status to a data asset.
- Data Literacy/Skills Assessment
 - This is an OMB mandate and due by July 31, 2020.
 - No additional information was discussed during the meeting.
- Broader Evidence Act Activities
 - The interim learning agenda activities are moving forward with three priority areas:
 - Drinking Water Systems Out of Compliance.
 - Grant Commitments Met.
 - Workforce Succession Management.
 - The Governance Council will be involved in some of these activities to help identify data sets available and potential data gaps to address questions that arise.

Data Maturity Model

- The CDO introduced the data maturity model activity.
 - EPA will need to select a maturity model to assess how it uses, manages, and governs its data.
 - Maturity models typically address how well an organization is doing now and the steps it needs to take to get the organization to where it wants to be.
- Office of Research and Development (ORD)/Office of Science Information Management (OSIM), provided an overview of ORD/OSIM's experience having Gartner conduct a benchmark study of how well ORD/OSIM is doing against similar research entities. This type of study is something EPA will need to conduct.

Subcommittees

- The CDO proposed the following subcommittees to organize the Council's priority activities. He will distribute a poll to help identify membership.
 - Inventory – How to best make the inventory complete and robust.
 - Data Literacy/Skills Assessment
 - Data Maturity Assessment

- Stewardship – How to bring the data stewards together to share best practices amongst themselves.
- Governance Structure – How to best govern EPA’s data.

Action Items

- Add the Data Lake as a potential Data Governance Council subcommittee.
- Governance Council members will inform the CDO of any COVID-19-related data assets their organization owns.
- Council members will respond to a poll to identify in which subcommittees they may be interested in participating. The poll will be distributed soon.