

EPA Data Governance Council

Meeting Summary – August 30, 2021

General Updates

Review the status of ongoing activities on the DGC Updates tab of the DGC Teams site.

- Metadata COTS Analysis
 - The Metadata COTS Analysis Subgroup completed testing of Collibra and Socrata.
 - GDIT is summarizing the findings from the hands-on testing and will deliver an update in about 2-3 weeks. The feedback will help to inform the Subcommittee's next steps.
- Data Maturity Assessment
 - The Data Maturity Subcommittee will be using Gartner's tool to conduct the maturity assessment.
 - The Subcommittee is working on communications to Deputy Assistant Administrators (DAAs) to explain the approach for the assessment and how it overlaps with other required activities under the Evidence Act.
 - The Subcommittee is also working on communications to EPA Information Management Officers (IMOs), Data Governance Council (DGC) members, and Senior Information Officials (SIOs) to request assistance with identifying individuals to take the self-assessments.
- Data Inventory
 - The Data Inventory Subcommittee is doing an analysis of EPA systems that contain data assets which will support the Environmental Dataset Gateway (EDG).
 - The Subcommittee is using data from READ to characterize the data assets.

Data Governance Strategy

- A new Subcommittee is being organized to focus on Governance and Strategy.
- The Subcommittee will hold bi-weekly meetings.
- Heather Barnhart will send the draft workplan to DGC members with the roster of current members. DGC members will identify any additional individuals from their offices who would like to participate on the Subcommittee.
- One of the Subgroup's first tasks will be to perform a Strength, Weakness, Opportunities and Threats (SWOT) analysis to help frame the development of the strategy.
- DGC Members provided initial thoughts on the SWOT analysis as follows:
 - Strengths
 - EPA has a significant amount of data and is making progress to improve its infrastructure to facilitate access to Agency data.
 - Automating data submission processes, such as Toxic Substances Control Act (TSCA) reporting via the Central Data Exchange (CDX), and eliminating paper-based processes has improved data quality and process efficiency.
 - Agency leadership is committed to identifying opportunities to expand and improve open data initiatives and values data as an Agency asset.

- Data owners are motivated to improve the quality of their data.
 - EPA's Geospatial Program is a model for best practices, methodologies, and tools to improve other Agency data management initiatives.
 - EPA has a data inventory/catalog, which many other agencies do not have.
 - EPA's Enterprise Information Management Policy (EIMP) provides a strong policy/legal framework to support open government initiatives.
 - Weaknesses
 - EPA does not have a deep understanding of how the public uses its data or what the Agency can do to make its data more valuable to the public.
 - EPA needs to define the different audiences (businesses, academia, etc.) that consume EPA's public-facing data.
 - EPA has multiple platforms and venues for exchanging data, which can complicate data integration and access.
 - EPA data is spread across organizational silos, which complicates integration.
 - Data management functions are not always perceived as a core responsibility.
 - EPA does not have a comprehensive understanding of how its data is used to support business functions.
 - Reliance on data from external sources (states, tribes, etc.) limits EPA's control over data quality and other data management functions.
 - EPA systems support many different business/stakeholder needs. Reaching consensus on requirements can be difficult.
 - The ability to waive or opt out of governance processes presents challenges to compliance with policies and consistent process implementation.
 - Quality Assurance (QA)/Quality Control (QC) processes and data management processes occur in parallel leading to potential communication gaps.
 - Opportunities
 - Identifying commonalities among Agency data management processes presents an opportunity to reduce duplication of effort.
 - Reviewing legacy system architectures can help to identify opportunities to improve interoperability (feeding data from transactional systems to support querying, etc.).
 - Improving communications about EPA's EIMP may improve understanding of Agency policy and drive adoption of open government initiatives.
 - Identifying overlap between QA/QC and data management processes could lead to opportunities for collaboration.
- Next Steps:
 - Heather Barnhart will compile input on the SWOT analysis for DGC and Subcommittee members to review.
 - Follow-on discussions will focus on refining the Opportunities and Threats.