

EPA Data Governance Council

Meeting Summary – March 28, 2022

General Updates

- EPA's Chief Data Officer, Richard Allen, welcomed the group.
- The DGC is spinning up new DGC subcommittees on Artificial Intelligence, Communications, CRADA, Data Skills Training and Metadata Stewardship.
- FY 22 ITPR Data Discussions will begin in April 2022. The two primary topics that will be discussed are managing unstructured data and the proposed data governance structure.
- Ongoing DGC project updates are available in the 03/28/2022 DGC Project Updates wiki.

Lanes of Opportunity (LoOps)

- Kristen Callahan presented an update on the LoOps.
- There are eight LoOps identified from the FY21 data discussions that address data challenges and opportunities.
 - Data Hub - An Informatics Integrateable Ready Data Repository
 - EPA Gateway powered by Login.gov
 - Modernization of the OARM Datamart
 - Data Governance Council Subcommittee on Governance and Strategy
 - Regulated Facilities Data Needs Project – Phase 2
 - Underground Injection Control (UIC) Data Application
 - CDM Phase 2 - PRIV / Privileged Access Management (PAM)
 - DLP for Network
- An overarching goal of the LoOps is to increase communications across EPA on some of our most important cross-cutting activities around data. The goal is to create more connectivity, share best practices and find new connections.

LoOps Deep Dive Presentation: Data Hub – An Informatics Integrateable Ready Data Repository

- Amar Singh presented an overview of the Data Hub project.
- EPA's Office of Research and Development manages research data through:
 - Development of QAPPs
 - Scientific data management plans for research activities
 - Use of ScienceHub to store data/metadata related to completed or published research
 - Creation of databases and systems to manage data
 - Data pipelines
- Data Hub was developed in an effort to implement a more comprehensive and coherent scientific information management strategy to increase research transparency and accessibility.
- Data Hub is a data repository which can be integrated for various research programs and application needs. It is designed to scale up in sustainable technology.

- The vision for Data Hub is to create data, information and knowledge management infrastructure and tools that can support efficient integration, transformation, dissemination, and mining of large-scale data on exposures and hazards related to environmental stressors and ecological integrity that provides:
 - Nimbleness for scientists to innovate while serving global outreach
 - Ability to connect data from anywhere
 - Enables inter-enterprise data sharing