



# Open Data in Biomedical Science: Policy Drivers and Recent Progress

*Ron Hines, PhD, ATS*  
*Associate Director for Health*



# Disclaimer and COI Statement

The views expressed in this lecture are those of the presenter and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

The presenter has no conflicts of interest to declare relative to the materials presented.



# Attribution

Jerry Blancato, Director, OSIM

Alex Montilla, OSIM

BJ George, NHEERL

Regrets from Allen Dearry



# Overview

- Policy drivers for open data in the federal government
- Open government initiative
- Existing open EPA data sources
- Implementation of Open Data directive at US EPA for data associated with publications





# Policy Drivers



4685

**Federal Register**  
Vol. 74, No. 15

## Presidential Documents

Monday, January 26, 2009

Memorandum of January 21, 2009

### Transparency and Open Government

Memorandum for the Heads of Executive Departments and Agencies

My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of **transparency, public participation, and collaboration**. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.....

A handwritten signature in black ink, appearing to be "Barack Obama", written in a cursive style.



# EPA's Open Government Plan 4.0

- Initial plan developed April 2010 in response to Presidential memorandum on Transparency and Open Government (FR 74.15, p. 4685, 2009)
- Final plan September 2016
- Underlying philosophy
  - Data must be managed as an asset and made available, discoverable and usable
  - “The better the Agency can inform the public and major stakeholders on what, how, and why the Agency is performing, the better it can work together with its major partners, such as the states, tribes, other agencies, and the general public, to meet our common goals.”



# EPA's Open Government Plan 4.0

- Objectives relative to open data
  - Catalogue and label information with metadata
  - Develop and maintain information using established standards and in machine readable formats
  - Make and maintain information to be open and publicly accessible unless covered by National Security Act or Controlled Unclassified Information Requirements
  - Increase the ability of internal and external stakeholders to analyze accessible data
  - Increase access to the results of EPA-funded scientific research
  - Adopt open source practices and repositories
  - Increase public access to custom software code where appropriate



# Existing EPA Open Data Sources

- Toxics Release Inventory (TLR)– 1986
  - Information on toxic chemical releases and other waste management data
  - Generate reports/maps on releases, transfers, and waste management, displaying by
    - Facility
    - Chemical
    - Geographic area
    - Industry (North American Industry Classification System code)
    - Reporting years
- EnviroFacts – 1995
  - Data about air, land, water, waste, toxics, radiation, regulated facilities, compliance grants, and regulated substances
  - Works in conjunction with MyEnvironment mapping tool



# Existing EPA Open Data Sources

- Health and Environmental Research Online (HERO)
  - >600K scientific references and data from the peer-reviewed literature used for Integrated Science Assessments (ISA), Provisional Peer Reviewed Toxicity Values (PPRTV); and Integrated Risk Information System (IRIS)
- Aggregated Computational Toxicology Resource (ACTOR)
  - Database on >500K chemicals from >1000 public sources
  - Searchable by chemical name (and other identifiers) and chemical structure
- ToxCast Dashboard
  - Data on >9K chemicals and information from more than 1,000 high-throughput assay endpoints
- EDSP21 Dashboard
  - Access to new, high throughput data relative to endocrine disruption on over 1,800 chemicals of interest



# Existing EPA Open Data Sources

- EnviroAtlas
  - Interactive resources for exploring the benefits people receive from nature or "ecosystem goods and services" (EGS)
  - Broad scale data for the lower 48 states and fine scale data for selected communities
  - Downloadable GIS toolboxes and ArcMap extensions that work with user-supplied data





# Policy Drivers

EXECUTIVE OFFICE OF THE PRESIDENT  
**OFFICE OF SCIENCE AND TECHNOLOGY POLICY**  
WASHINGTON, D.C. 20562

February 22, 2013

**MEMORANDUM FOR THE HEADS OF EXECUTIVE OFFICES AND AGENCIES**

**FROM:** John P. Holdren   
Director

**SUBJECT:** Increasing Access to the Results of Federally Funded Scientific Research



# Policy Drivers

## Holdren Memo

- Commitment to ensuring that the direct results of federally funded research are made available and useful to the public, industry and the scientific community
  - Fewest constraints possible
  - Consistent with law
  - Includes peer-reviewed publications and digital data
- Recognized services provided by publishers, *e.g.*, peer review coordination
  - Important for the high quality and integrity of scholarly publications
  - Must continue to be made available
  - 12 month embargo on making published articles available
- Federal agencies with > \$100 million in annual R&D expenditures must develop plan to support increased public access to research results



# Implementation Progress at EPA

- Development of tools and implementation plan by EPA/ORD's Office of Scientific Information Management (OSIM)
- Two tract system, one for actual manuscripts and a second for data sets
  - ORD subscribed to NIH PubMed Central FY16 Q3 as a repository for ORD intramural research publications
    - 12 month embargo period consistent with Holdren Memo recommendations
  - Batch submission of manuscripts from EPA's **Scientific and Technical Information Clearance System** (STICS)
  - ORD Scientific Data Management Plan standards extended to comply with open data requirements of Holdren Memo
    - Does not cover research data shared with EPA but owned by other organizations
    - Release within 30 days of publication being deposited into PubMed Central



# Implementation Progress at EPA

- Scientific Data Management Plan standards
  - Describe all collected or created research data and *metadata*
  - Plans for long-term preservation of, and access to research data, as appropriate
  - Responsibility to protect confidentiality and personal privacy, respect proprietary interests and property rights
  - At a minimum, research data associated with and supporting conclusions of any manuscript to be shared
  - Describe location, primary contacts, organization, and access restrictions
  - Data security included



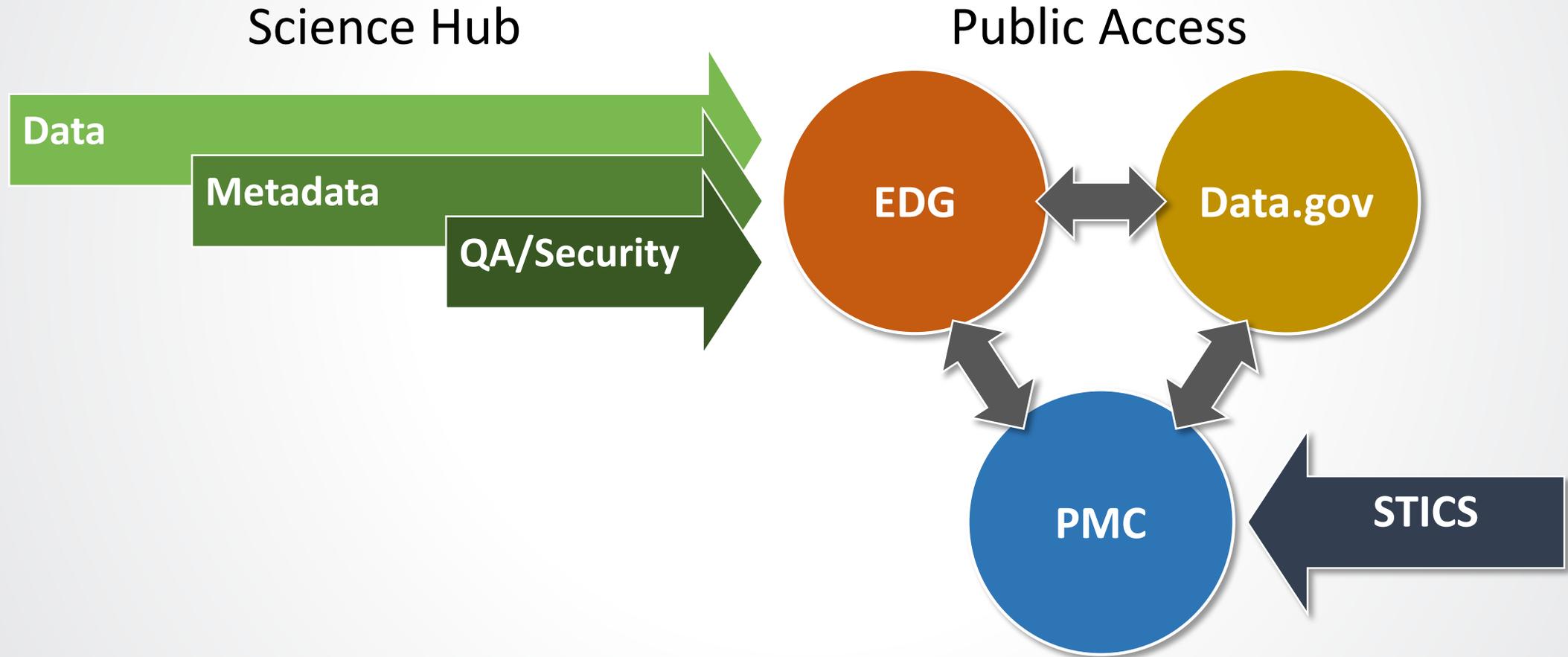
# Implementation Progress at EPA

- Metadata
  - Descriptive information about data sufficient to make data independently understandable to someone not involved in the research effort
  - Fields pointing to both the data repository and web services and/or application programming interfaces available to access the data
- Cross-referencing data sets and manuscripts
  - Persistent URL assigned to both manuscript and dataset and referenced in both
  - Unique Digital Object Identifier (DOI) assigned to both manuscript and dataset and referenced in both
  - Reference to lead investigator ORCID number included in both manuscript and data set



# Implementation Progress at EPA

- ScienceHub developed by OSIM as EPA's dataset clearance system
  - Repository and clearance system for EPA's data management plans and datasets
  - Coordination with EPA's Quality Assurance Program
  - Once data cleared in ScienceHub, publication in:
    - Environmental Data Gateway– home of EPA's open data
    - Data.gov – home of the federal government's open data
  - As of Feb 24, 2017, 134 datasets deposited into ScienceHub and 45 data sets published on the Environmental Data Gateway and Data.gov
- ScienceHub Phase II
  - Enable investigator teams to manage and share data for an entire research effort
  - Once datasets finalized and cleared, made publicly accessible via the Environmental Data Gateway and Data.gov





# Questions

