

# The Evolution of Data.gov

---





## MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Transparency and Open Government

*Government should be transparent.* Transparency promotes accountability and provides information for citizens about what their Government is doing. Information maintained by the Federal Government is a national asset. My Administration will take appropriate action, consistent with law and policy, to disclose information rapidly in forms that the public can readily find and use. Executive departments and agencies should harness new technologies to put information about their operations and decisions online and readily available to the public. Executive departments and agencies should also solicit public feedback to identify information of greatest use to the public.

**“Information  
maintained by the  
Federal  
Government is  
a national asset.”**

*President Obama  
January 21, 2009*

# Data.gov Conceptual Underpinnings

- ◆ The free flow of information between the government and the public is essential to a democratic society
- ◆ Increasing the ability of the public to discover, understand, and use the wealth of government data increases government accountability and unlocks additional economic and social value.

# Two Levels of Data.gov

- Policy Level
  - Increasing access
    - Finding data that are currently available
    - Making additional data sources available
      - Opening new sources
      - Removing fees and registrations
  - Increasing usability
    - Formats
    - Fitness for use data
    - Machine readable metadata
    - Search capabilities
    - Analytic capabilities
- Platform
  - Web site is public presence
  - Demonstrating Administration's commitment to transparency
  - Concrete first step – the current site is the initial base which will evolve

# Finding and Using Government Data

There are over 24,000 .gov web sites – difficult to navigate

Currently the public must understand the Federal organizational structure to find datasets on key topics

Data are not always  
downloadable for analysis;  
sometimes locked in legacy systems

Web sites not designed for new generation  
of technologies on the web

# Objectives of the Data.gov Platform (the website)

- ◆ Transcend stovepipes to increase access to data
  - ◆ Cross-government metadata catalog
  - ◆ Downloadable data only one click away
- ◆ Encourage development of innovative applications by presenting data in formats that can be analyzed, visualized, and mashed-up
- ◆ Provide ability for user to determine fitness of a dataset for a particular use
- ◆ Provide tools that allow access to complex data stores, real-time data (RSS feeds), and downloadable applications (widgets)

# Evolving Scope of Data.gov

- ◆ Machine readable datasets that:
  - ◆ Are already in the public domain (web accessible)
  - ◆ Are hosted by the Executive Branch of the Federal Government
  - ◆ Do not have copyright limitations
  - ◆ Do not impose charges for accessing data
  - ◆ Do not require registration to access data
  - ◆ Support the preference for “raw” (usable) data

# What is meant by “raw” data?

- ◆ Machine readable
- ◆ Preference for structured datasets
- ◆ Data that can be used (analyzed) for multiple purposes
- ◆ Preference for the most granular data that meet applicable privacy and security protections
- ◆ Datasets that can be mashed-up with others (brought into context using lightweight web programming techniques and technologies)

# Preferred Formats for Datasets

## ◆ Preferred formats for “raw” data:

- XML
- CSV/TXT
- RSS
- Keyhole Markup Language (KML/KMZ)
- ESRI Shapefile

## ◆ Avoid datasets in the following formats:

- PDF
- HTML (e.g., HTML tables containing data)

# Data Policy

<http://www.data.gov/datapolicy>

1. **Public Information:** Datasets will not contain National Security information as defined by statute and/or Executive Order, or other information/data that are protected by other statute, practice, or legal precedent.
2. **Security:** Agencies are required to ensure datasets are compliant with required confidentiality, integrity, and availability controls mandated by FIPS 199.
3. **Privacy and Confidentiality:** Agencies are required to ensure datasets are compliant with current privacy and confidentiality requirements.
4. **Data Quality :** All information accessed through Data.gov is subject to the Information Quality Act (P.L. 106-554). For all data accessed through Data.gov, each agency has confirmed that the data being exposed through this site meet the agency's Information Quality Guidelines.

# Agency/Program Retains Control of its Data

- ◆ The dataset continues to reside on agency's own website
- ◆ Agency retains responsibility for authoritative source of data, including quality, corrections, and updates
- ◆ Agency determines datasets to submit based on:
  - Data policy, including :
    - Information quality
    - PII and potential mash-ups
  - Exposure implications
    - Ready to support questions on the dataset
    - Agency site performance
- ◆ Agency provides metadata

Featured Dataset or Tool

Search by State or local government

Suggest Datasets

FEATURED TOOL:

## BUREAU OF TRANSPORTATION STATISTICS Airline On-Time Performance and Causes of Flight Delays

The Airline On-Time Performance and Causes of Flight Delays table contains on-time arrival data for non-stop domestic flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance.

[VIEW THIS TOOL](#) ▶



Search the following Data.gov catalogs:

CSV  
XML  
KML  
SHP

"RAW" DATA CATALOG

TOOL CATALOG

GEODATA CATALOG

Search Catalogs

### Welcome to Data.gov

The purpose of Data.gov is to increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government. Although the initial launch of Data.gov provides a limited portion of the rich variety of Federal datasets presently available, we invite you to actively participate in shaping the future of Data.gov by suggesting additional datasets and site enhancements to provide seamless access and use of your Federal data. Visit today with us, but come back often. With your help, Data.gov will continue to grow and change in the weeks, months, and years ahead.

### How to use Data.gov

Data.gov includes searchable [data catalogs](#) providing access to data in three ways: through the "raw" data catalog, the tool catalog and the geodata catalog. Please note that by accessing datasets or tools offered on Data.gov, you agree to the [Data Policy](#), which you should read before accessing any dataset or tool. If there are additional datasets that you would like to see included on this site, [please click here](#). For more information on how to use Data.gov, [view our tutorial](#).

Data.gov Tutorial



# Easily Downloadable Data: A Click Away

2007 Toxics Release Inventory National data file of all US States and Territories

**DATASET SUMMARY**

<b>Agency:</b>	Environmental Protection Agency
<b>Category:</b>	Geography and Environment
<b>Date Released:</b>	3/19/2009
<b>Date Updated:</b>	6/10/2009
<b>Time Period:</b>	Calendar Year 2007
<b>Frequency:</b>	Annual
<b>Description:</b>	The Toxics Release Inventory (TRI) is a publicly available information on toxic chemical releases and waste management annually by certain industries as well as federal facilities.

**DATASET RATINGS**

	Current	Your Rating [?]
Overall	★★★★★ (3 votes)	★★★★★
Data Utility	★★★★★ (3 votes)	★★★★★
Usefulness	★★★★★ (3 votes)	★★★★★
Ease of Access	★★★★★ (3 votes)	★★★★★

**DATASET INFORMATION**

**DOWNLOAD INFORMATION**

Feeds ▶	XML ▶
CSV/TXT ▶	XLS ▶
<b>CSV</b> 37.40MB	
KML/KMZ ▶	Shapefile ▶
Maps	

**Comment on these data:**

ritzy 8:30

Type the two words:

**SUBMIT**

(Privacy Policy)

**Download data here**

**Rate data here**

**Comment here**



# Senior Advisory Group

- ◆ Provides OMB with a forum for working with those responsible for data generation and dissemination
- ◆ Provides advice to OMB regarding the strategic vision and transformative potential for Data.gov
- ◆ Provides feedback to OMB on the potential impact of Data.gov proposals on agency and interagency data generation and dissemination programs
- ◆ Encourages the development and implementation of a unified vision for achieving data interoperability and other efforts to modernize Federal data dissemination and sharing

# Initial Communities of Interest on the Senior Advisory Group

- ◆ Inter-agency efforts that focus on missions associated with data policy, information management, and enhancing the dissemination of digital data
- ◆ Examples of these formal communities of interest that address data policy within the Executive Branch include:
  - The Council of Chief Information Officers (CIO Council)
  - The Interagency Council on Statistical Policy (ICSP)
  - The Federal Geographic Data Committee (FGDC)
  - CENDI (senior scientific and technical information (STI) managers)
  - The Interagency Working Group on Digital Data
  - The Networking and Information Technology Research and Development (NITRD) Program

# Contact Information

Paul Bugg

202-395-3095

[pbugg@omb.eop.gov](mailto:pbugg@omb.eop.gov)

Margo Schwab

202-395-5647

[mschwab@omb.eop.gov](mailto:mschwab@omb.eop.gov)

# Data.gov

---

Questions?