

# Opening the Door to Government Data

Records Administration Conference  
May 12, 2010

Jim Rolfes, Program Director, Data.gov and Chief  
Enterprise Architect, U.S. Department of the Interior

# Launched on May 21, 2009, Data.gov is a flagship of the Administration's Open Government Agenda

As of May 4, 2010:

- 1,284 Raw Datasets
- 495 Tools
- 167,952 Geodata
- **169,731 Total**

The screenshot shows the Data.gov website interface. At the top, it displays the date 'TUESDAY, MAY 04, 2010' and the text 'AN OFFICIAL WEB SITE OF THE UNITED STATES GOVERNMENT'. The main header features the 'DATA.GOV' logo and social media sharing options. A navigation menu includes links for HOME, OPEN GOVERNMENT, FEDERAL, STATE/LOCAL/TRIBAL, SUGGEST DATASETS, DIALOGUE, METRICS, FAQ, and ABOUT. The main content area is split into two sections. On the left, a blue banner with the text 'DISCOVER. PARTICIPATE. ENGAGE.' promotes the catalogs, with icons for 'RAW DATA CATALOG' (CSV, XML, KML, SHP), 'TOOL CATALOG' (wrench and screwdriver), and 'GEODATA CATALOG' (globe). On the right, a featured dataset titled 'City and County Web Dataset' is shown with a red pushpin on a map and a description: 'The geographic names dataset provides a "mashup" of URLs for city and county web sites and city and county location data from the USGS Geographic Names Information System (GNIS). GNIS data includes incorporated places, census designated areas, unincorporated places, counties, and populated places.' A 'View more' button is visible below the description. At the bottom of the page, there is a search bar with a 'SEARCH' button and a link to 'ADVANCED SEARCH', and a section titled '★ New Integrated Search' with a brief description of the partnership with Search.U.S.A.gov.

## 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

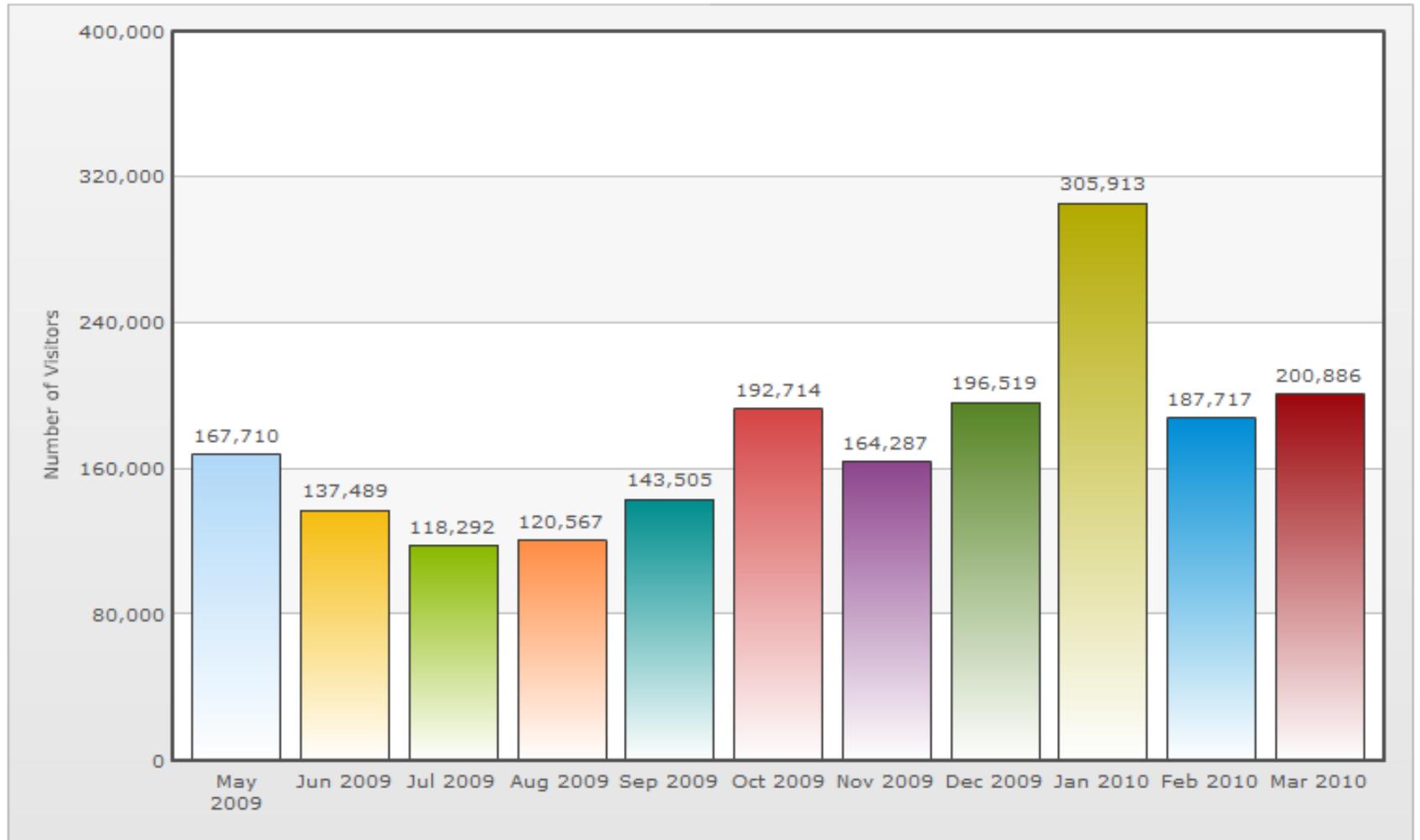
## ★ New Integrated Search

Data.gov has partnered with **Search.U.S.A.gov** to optimize our website and to take advantage of their

# Year to Date Number of Visitors: 1,935,599

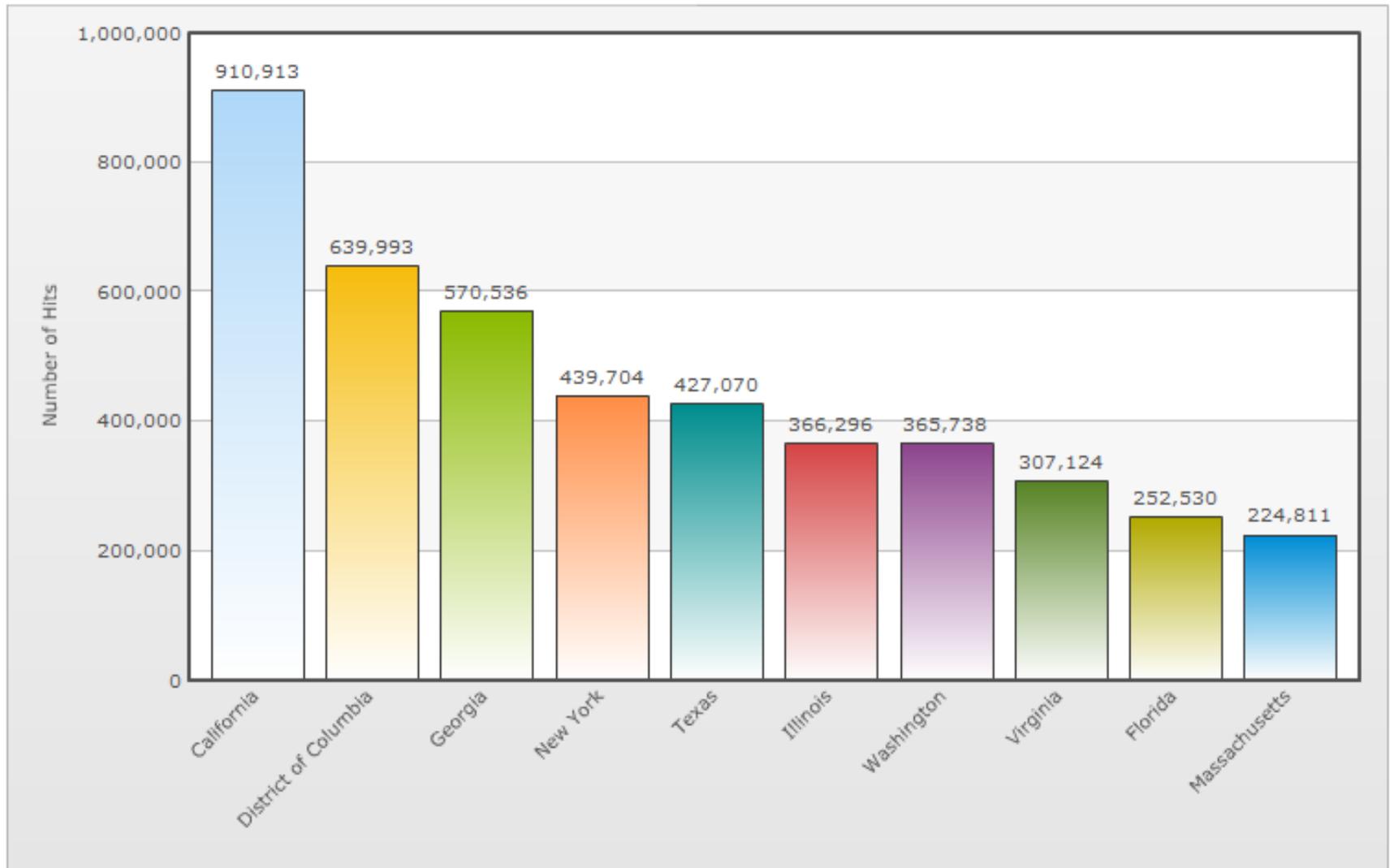
## Data.gov Monthly Visitor Statistics

Year to Date Number of Visitors : 1,935,599



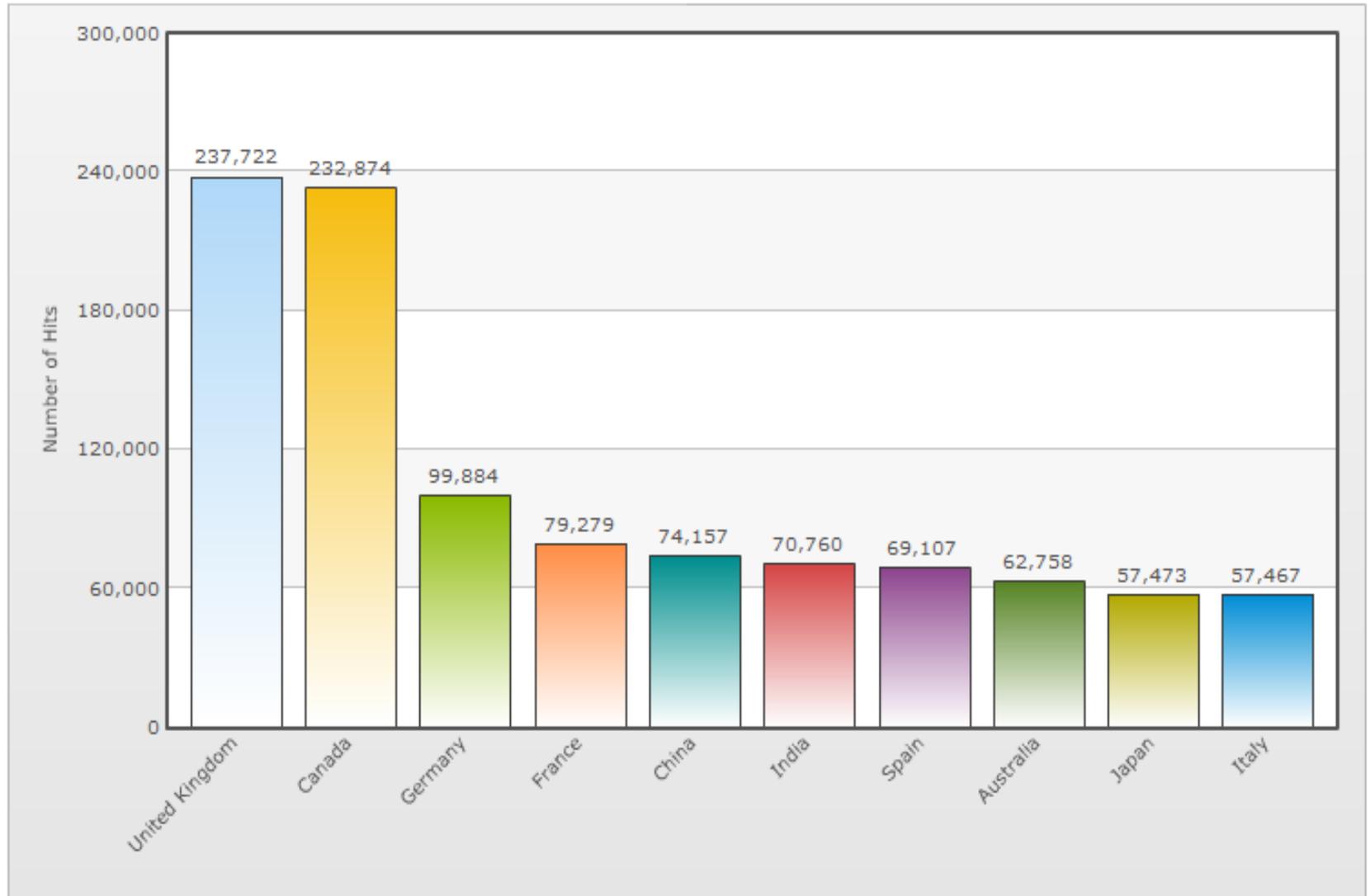
# Top 10 Visiting States (March 2010)

Top 10 Visiting States (March 2010)



# Top 10 Visiting Countries (March 2010)

Top 10 Visiting Countries (March 2010)



# Goals of Data.Gov

## FEATURED TOOL:

### U.S. GEOLOGICAL SURVEY (USGS)

#### USGS Global Visualization Viewer for Aerial and Satellite Data

Ten million archive images of the Earth's surface are available for immediate selection and free download via the USGS Earth Resources Observation and Science (EROS) Center's Global Visualization Viewer at <http://glovis.usgs.gov/>. Users can preview thumbnails, browse images and download full-image selections from 1.5 million aerial photos of U.S. sites and 8.5 million images captured worldwide by U.S. Earth-observing satellites.



[VIEW THIS TOOL](#) ▶

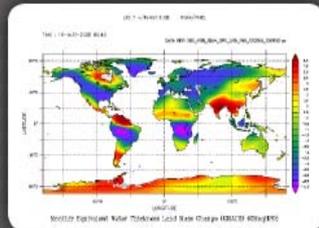
- Break down stovepipes to **increase access** to data
- Instant access: Downloadable data **only one click away**
- Data that can be **manipulated and mashed-up**
- Encouraging development of **innovative applications**
- Tools that **provide access** to complex data stores
- Widgets

## FEATURED TOOL:

### NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA)

#### MyNASA

Mentoring and inquiry using NASA data on Atmospheric and earth science for Teachers and Amateurs. The MY NASA DATA Live Access Server (LAS) is now available to create your own microsets for your class or your interests. The LAS contains over 149 parameters in atmospheric and earth science from five NASA scientific projects.



[VIEW THIS TOOL](#) ▶

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:



Search Catalogs

### Welcome to Data.gov

The purpose of Data.gov is to increase public access to 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

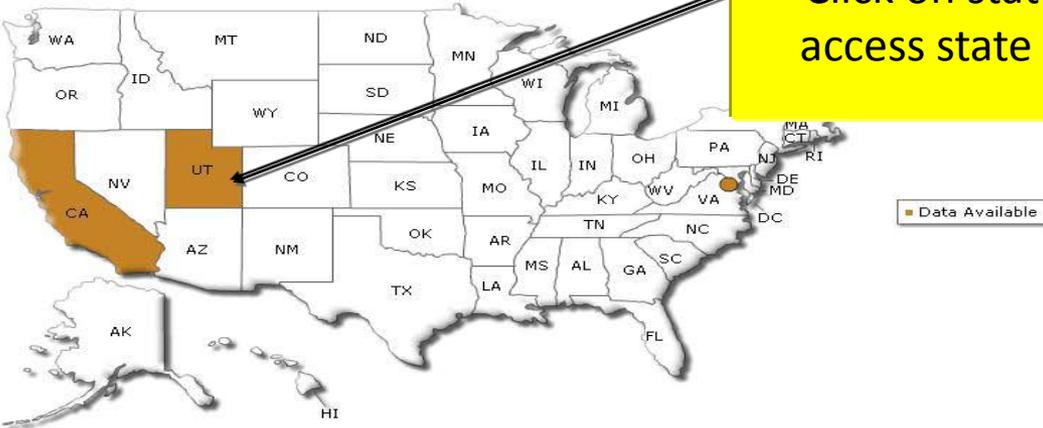
Data.gov includes searchable **data catalogs** providing access to data in three ways: through the "raw" data catalog, the tool catalog and the geodata catalog. Please note that by accessing datasets or tools offered on Data.gov, you agree to the **Data Policy**, which you should read before accessing 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

# Engaging a Larger Community- State and Local Data

## State and Local Data Sites

We are encouraging and working with states and local governments to launch their own sites with access to machine-readable data. The map below provides easy access to the state and local data sites that have been launched.



# Catalogs

## Data.gov Catalogs

Use the Data.gov catalog below to access U.S. Federal Executive Order 13526, which requires the release of certain information. By accessing the data catalogs, you can view the "raw" data catalog, using tools and through the geodata catalog. The "raw" data catalog provides machine-readable, platform-independent datasets while the Tools Catalog provides tools that allow you to mine datasets.

Input keywords



Navigation tabs: "RAW" DATA CATALOG, TOOL CATALOG, GEODATA CATALOG

Search tools by keywords:

Search tools by single/multiple category:

- All Categories
- Births, Deaths, Marriages, and Divorces
- Business Enterprise

Search tools by single/multiple agency:

- All Agencies
- Agency for Healthcare Research and Quality (AHRQ)
- Bureau of Economic Analysis (BEA)

Page 1 of 2 (35 records found)

Navigation: << < 1 | 2 > >>

Results per page: 25 | 50 | 100

Name (click for metadata and to rate tool)	Rating	Agency	Category	Data Extraction	Widget
<b>Airline On-Time Performance and Causes of Flight Delays</b> This 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 a...	★★★★★ (5 votes)	BTS	Transportation		
<b>American FactFinder</b> American FactFinder is the U.S. Census Bureau's online, self-service tool designed to search the American Community Survey, Puerto Rico Community Survey, Decennial Census, ...	★★★★★ (12 votes)	Census	Population		
<b>Cancer Incidence - Surveillance, Epidemiology, and End Results (SEER) Registries Limited-Use</b> SEER Limited-Use cancer incidence data with associated population data. Geographic areas available are county and SEER registry. The Surveillance, Epidemiology, and End ...	★★★★★ (7 votes)	NCI	Health and Nutrition		
<b>CDC Emergency Text Messages</b> Provides up-to-date information on public health text alerts before and during emergencies.	★☆☆☆☆ (0 votes)	CDC	Health and Nutrition		
<b>DataFerrett</b> The DataFerrett tool is used to search American Community Survey Public Use Microdata, Current Population Survey(CPS), CPS supplemental surveys, Survey of Income and Pro...	★★★★★ (17 votes)	Census	Population		

Rate the tool

# Easily Downloadable Data

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 by certain industries as well as federal facilities.

DOWNLOAD INFORMATION	
<a href="#">Feeds &gt;</a>	<a href="#">XML &gt;</a>
<a href="#">CSV/TXT &gt;</a>	<a href="#">XLS &gt;</a>
<a href="#">CSV</a> 37.40MB	<a href="#">Shapefile &gt;</a>
<a href="#">XML/KML &gt;</a>	
<a href="#">Maps</a>	

Download data here

Comment on these data:

[SUBMIT](#)

[\(Privacy Policy\)](#)

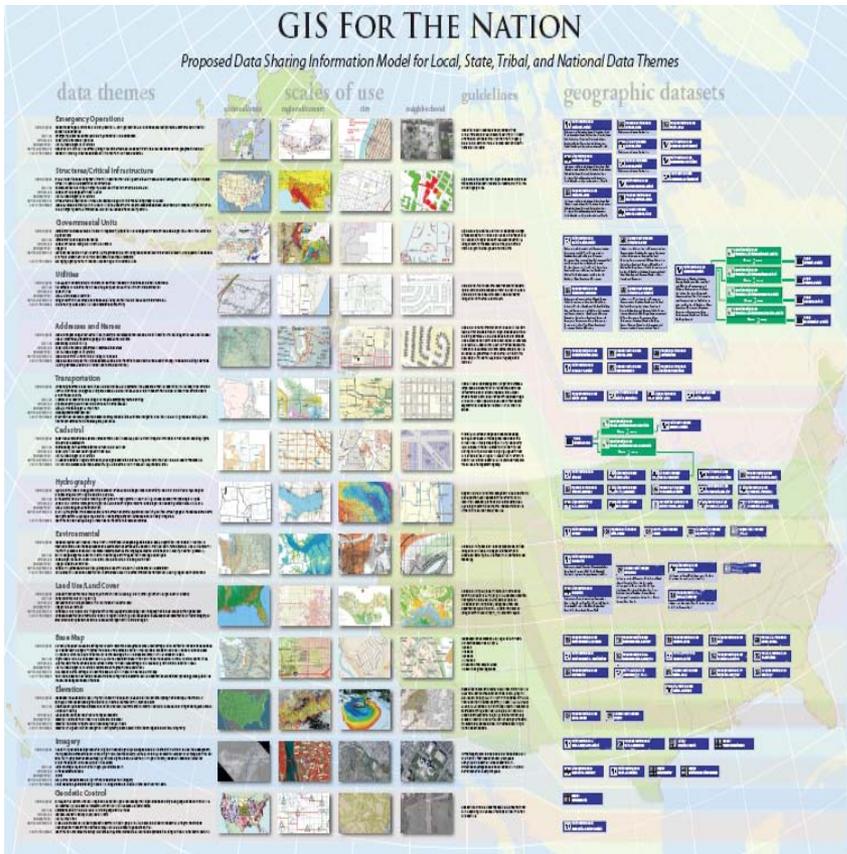
DATASET RATINGS	
	<b>Current</b> <b>Your Rating [?]</b>
<b>Overall</b>	★★★★★ (3 votes) Ⓞ ★★★★★
<b>Data Utility</b>	★★★★★ (3 votes) Ⓞ ★★★★★
<b>Usefulness</b>	★★★★★ (3 votes) Ⓞ ★★★★★
<b>Ease of Access</b>	★★★★★ (3 votes) Ⓞ ★★★★★

Rate data here

Comment here

DATASET INFORMATION	
<b>Data.gov Data Category Type</b>	Raw Data
<b>Specialized Data Category Designation</b>	Administrative
<b>Keywords</b>	TRI, TRI Data, TRI Reporting, Toxic, Toxics, Toxic Release, Toxics Release, Toxic Release Inventory, Toxics Release Inventory, Chemical, Chemicals, Chemical Release, Chemical Pollution, EPCRA, EPCRA 313, section 313, TRI State Data, Hazardous, Right to Know, Form R, Form A, pollution prevention, waste management, source reduction, community, community right to know, toxic chemical release inventory, toxic chemicals release inventory, facility, facilities

# GeoData is a key enabler



- “Human knowledge is expected to be doubling by the year 2012” (Alvin Toffler)
- Geo-aware Applications are key
- How To Realize the Promise of GIS for the Nation?
  - Geo-data “architected and built-in”
  - Lightweight Geo-Applications and Mash-ups for Web 2.0/Gov 2.0
  - Incremental, Agile, Actionable and Affordable Delivery

# Agency and Citizen Engagement is Key

Agencies Providing Data to Include on Data.gov and Exposing New Datasets



Citizens Using Data.gov Data to Create Applications and Mash-ups

Citizens Suggesting Datasets They Want To See

Citizens Commenting on Specific Datasets

DATASET RATINGS		
	Current	Your Rating [?]
Overall	★★★★★	☉ ★★★★★
Data Utility	★★★★★	☉ ★★★★★
Usefulness	★★★★★	☉ ★★★★★
Ease of Access	★★★★★	☉ ★★★★★

Data.gov Users Rating Datasets Using a Five Star Rating

# Privacy, Quality, and Records Management Considerations

- Datasets continue to reside on Agency's own website
- Agency retains responsibility for authoritative source of data, including corrections and updates
- Agency retains responsibility for protecting personally identifiable information, information quality, and records retention
- Agency retains responsibility for impacts of Data.gov access on their site

# Data.Gov Future...

- Continue to Tap into the ingenuity of the American People
- More collaboration and participation with the Public (e.g., blogs)
- Social interaction and discussion
- Implementation of new technologies (e.g., web services and RSS feeds)
- Additional datasets, including those not already publicly available
- More Data visualizations, Mash-ups and applications from the data on Data.gov
- Linked Open Data...Semantic Web etc.

# Resources

- NARA Toolkit for Managing Electronic Records:  
<http://www.archives.gov/records-mgmt/toolkit/#list>
- Data.Gov <http://www.data.gov>
- Open Government  
<http://www.whitehouse.gov/open>