Sanjeev "Sonny" Bhagowalia
CIO, U.S. Dept. Of The Interior

Linda A. Travers
Acting Assistant Administrator and CIO,
U.S. Environmental Protection Agency

Co-Leads Data.gov
Federal CIO Council
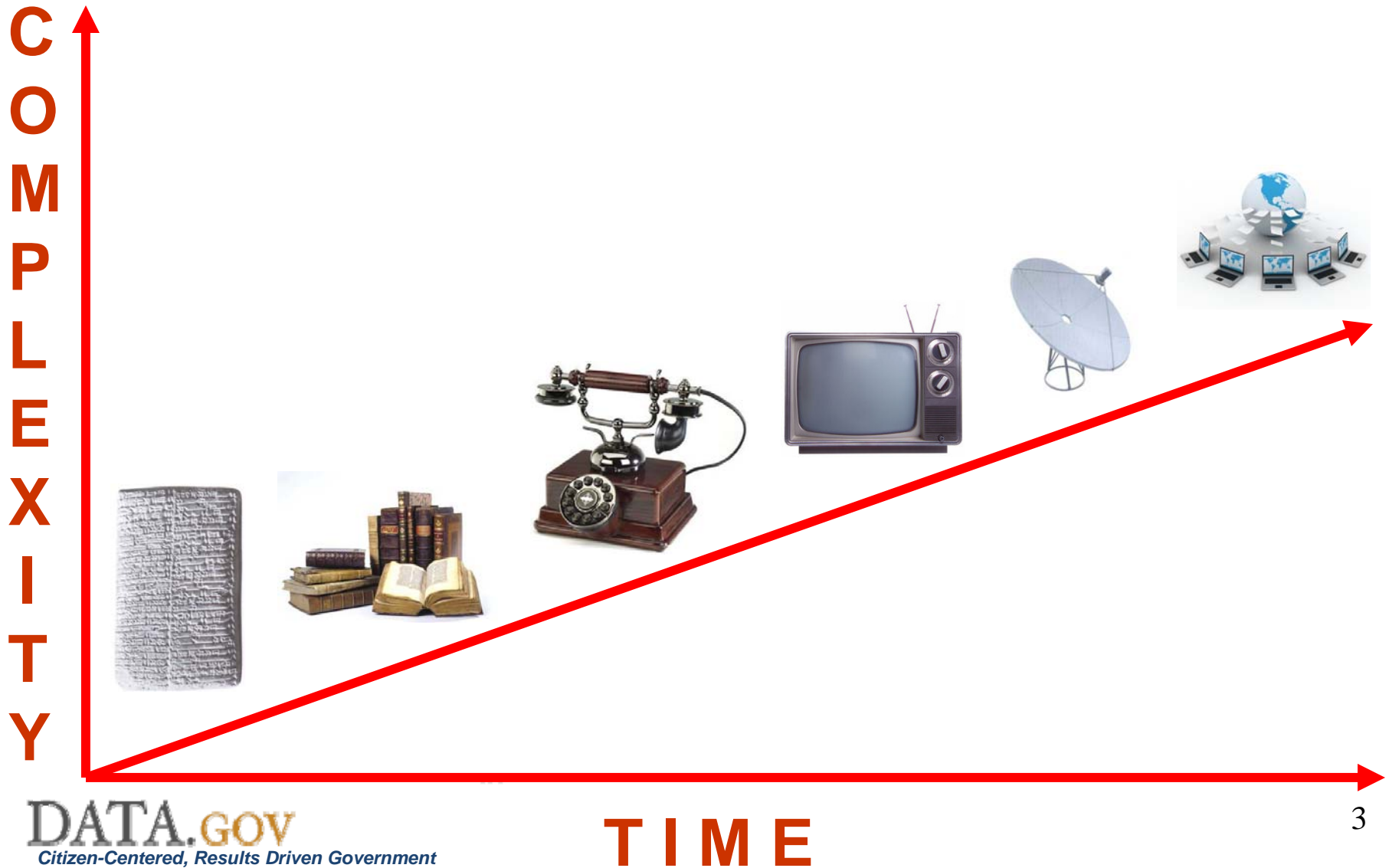
OMB

DOI

EPA

GSA

The Issues

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

Public must understand the federal organizational structure to find datasets on key topics

Data exists in portals – are not "raw" machine downloadable or data locked in legacy systems

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

The Communications Revolution

C O M P L E X I T Y

T I M E

DATA.GOV
Citizen-Centered, Results Driven Government

Rapid Increase In Technology

In just 60 years, technology has advanced so quickly our way of life has changed

We must continually evaluate how we share data and information

Unlocking the Wealth of Data is Key

DATA +
CONTEXT =
INFORMATION

Set the Information FREE!

the WHITE HOUSE PRESIDENT BARACK OBAMA ★ ★ ★ ★

the BRIEFING ROOM   the AGENDA   the ADMINISTRATION

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
Citizen-Centered, Results Driven Government

When Data is Democratized

When the Department of Defense released its satellite data…

the public <u>created</u> affordable GPS devices!

DATA.GOV
Citizen-Centered, Results Driven Government

Tapping the Ingenuity of the American Public

- The Federal Government does not have a monopoly on good ideas

- Access to Information unleashes creativity
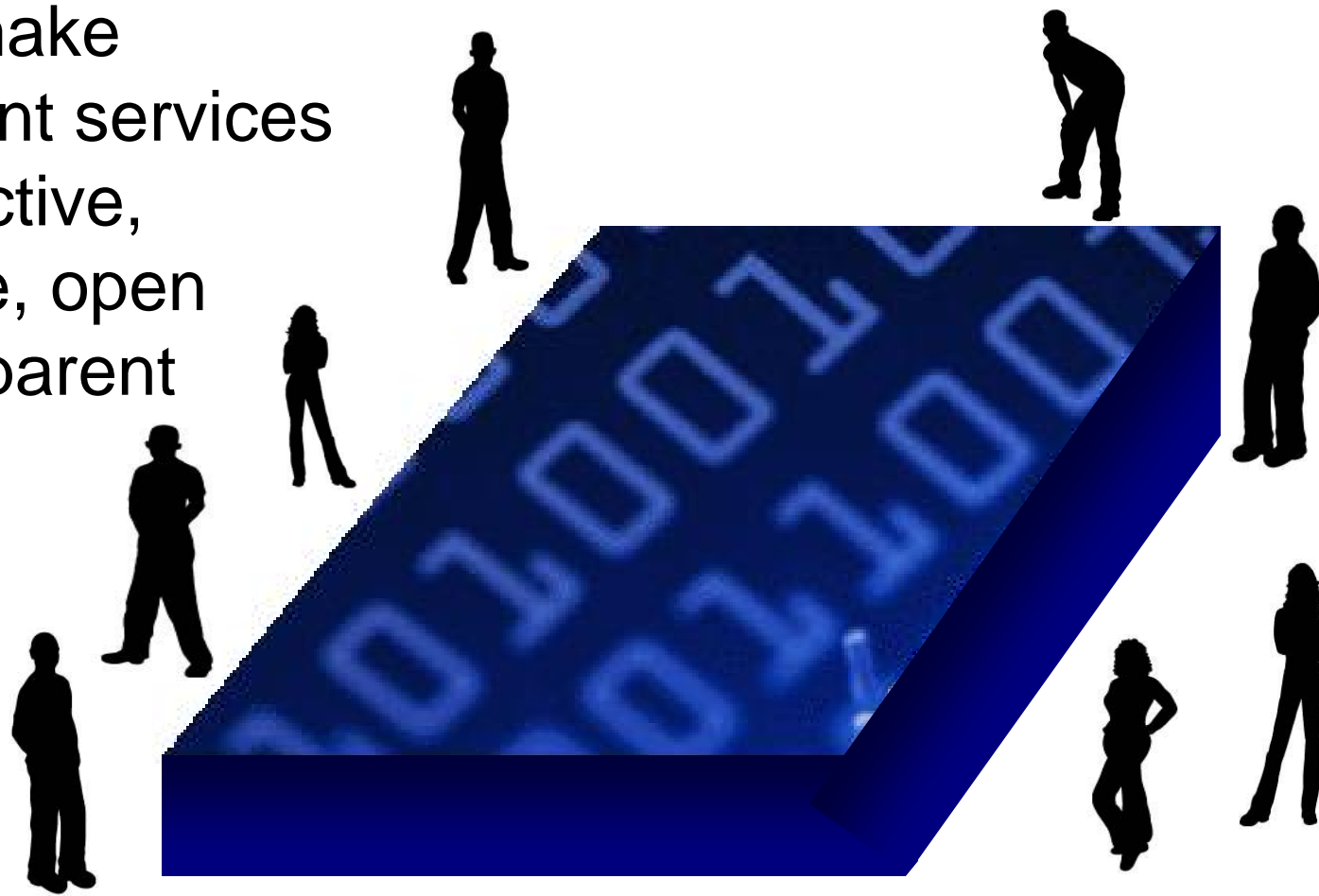
- Innovation drives sustainability

The Digital Public Square

A place for sharing ideas to make government services more effective, accessible, open and transparent

Part of President Obama's Open Government Initiative

- *A government that is:*
  - **Transparent**
    - Promotes accountability
    - Provides information for citizens on what their government is doing
  - **Participatory**
    - Agencies encouraged to provide citizens opportunities to participate in policy making
    - Agencies encouraged to solicit ideas from citizens about how to improve those opportunities
  - **Collaborative**
    - Use innovative tools to enable collaboration across and at all levels of government

DATA.GOV
Citizen-Centered, Results Driven Government

10

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.
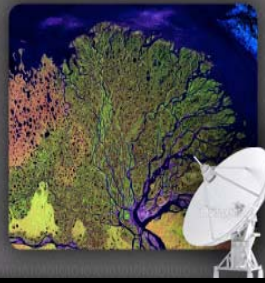
VIEW *THIS* TOOL ›

FEATURED TOOL:
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA)
MyNASA

Mentoring and inquiry using NASA data on Atomspheric 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.
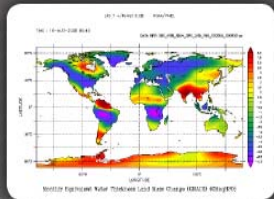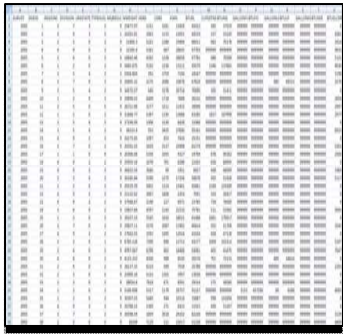
VIEW *THIS* TOOL ›

◆ **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 datastores**

◆ **Widgets**

DATA.GOV
Citizen-Centered, Results Driven Government

11

Data Quality and Privacy are Paramount

◈ The dataset continues to reside on agency's own web site

◈ 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
Citizen-Centered, Results Driven Government

What is Meant by 'Raw' Data?

Machine readable formats
- RSS
- CSV
- KML
- SHP
- XLS
- XML

◆ Machine readable

◆ Preference for structured datasets

◆ Data that can be used for multiple purposes

◆ Data that is largely free from interpretation and value judgments (observational, unbiased data)

◆ Datasets that can be mashed-up with others (brought into context using lightweight web programming techniques and technologies)

◆ As 'raw' as possible (lowest level of aggregation, before analysis)

◆ Preferred formats for 'raw' data: XML, CSV/TXT, RSS, Keyhole Markup Language (KML/KMZ), ESRI Shapefile

◆ Avoid datasets in the following formats: PDF and HTML (e.g., HTML tables containing data)

DATA.GOV
Citizen-Centered, Results Driven Government

DataGov went Live on May 21, 2009 - Leveraging Innovative, New Cloud Computing Technologies

47 Raw Data Sets and 27 Tools = 74 in Catalog (Start)

579 Raw Data Sets; 208 Tools and 109, 471 Geo Data = 110, 258 in Catalog (Today)

Users

Request/Response

"Cloud"   Content Delivery Network

Origin Load Balancer

Ranking/ Rating

Read Only

Master ← Web Servers ← Slave (DB Servers)

Sync
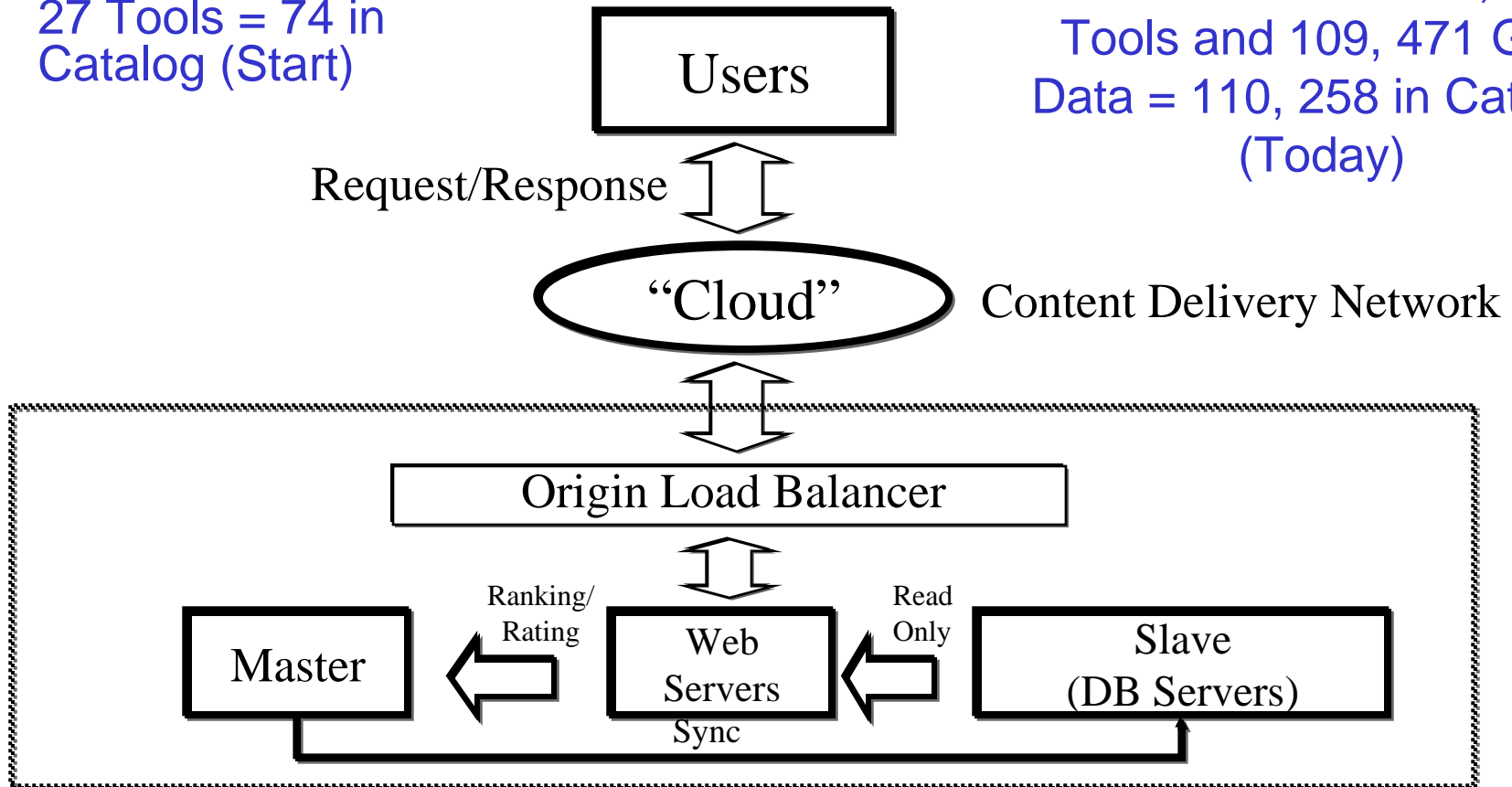
Over 23 Million hits since launch with 91.6% handled by Cloud!

14

Top Ten States Visiting Data.Gov

◈ California
◈ District of Columbia
◈ New York
◈ Texas
◈ Virginia
◈ Illinois
◈ Massachusetts
◈ Washington
◈ Florida
◈ Georgia

DATA.GOV
Citizen-Centered, Results Driven Government

15

Top Ten Countries Visiting Data.Gov

- USA
- Canada
- United Kingdom
- Germany
- Australia
- France
- Netherlands
- China
- Italy
- India

DATA.GOV
Citizen-Centered, Results Driven Government

DATA.GOV
Citizen-Centered, Results Driven Government

Engaging a Larger Community - State and Local Specific Data

Catalogs

Easily Downloadable Data: A Click Away

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

DATASET SUMMARY

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

DOWNLOAD INFORMATION

Feeds ▶   XML ▶

CSV/TXT ▶   XLS ▶

☰ CSV
37.40MB

KML/KMZ ▶   Shapefile ▶

Maps

Download data here

DATASET RATINGS

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

ritzy   B:30

Type the two words:

reCAPTCHA™
stop spam.
read books.

SUBMIT

Rate data here

Comment on these data:

SUBMIT

(Privacy Policy)

Comment here

DATASET INFORMATION

Data.gov Data Category Type	Raw Da...
Specialized Data Category Designation	Adminis...
Keywords	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

DATA.gov
Citizen-Centered, Results Driven Government

Geo-Spatial Data is A Key Enabler for "Killer Apps"

GIS FOR THE NATION
Proposed Data Sharing Information Model for Local, State, Tribal, and National Data Themes
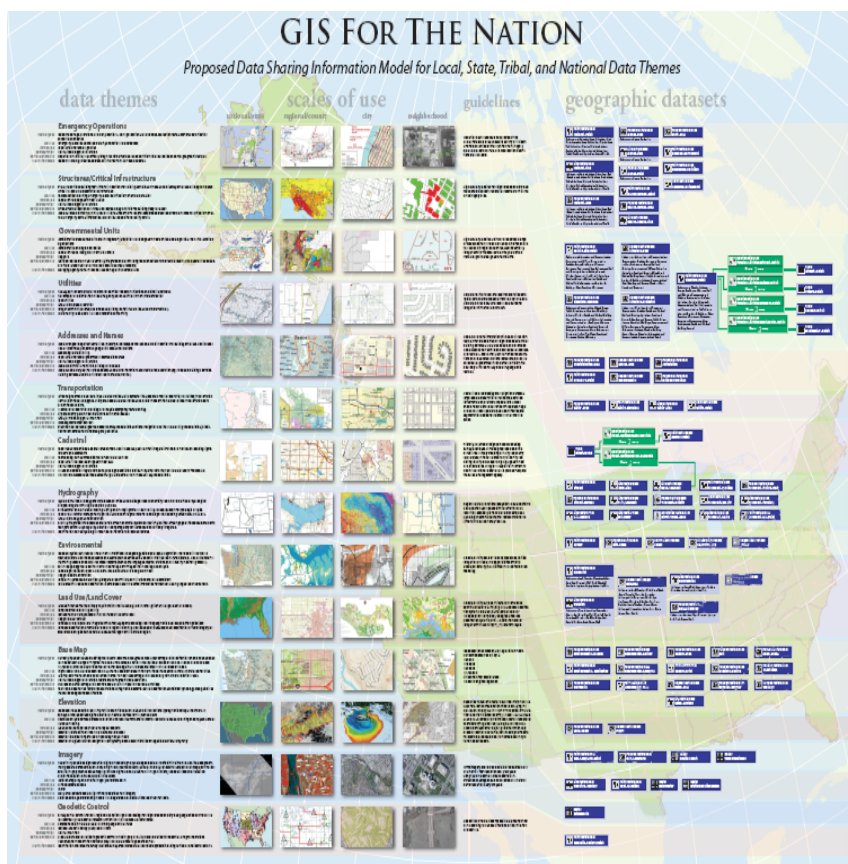
- ◆ "Human knowledge is expected to be doubling by the year 2012" (Alvin Toffler)

- ◆ 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

DATA.GOV
Citizen-Centered, Results Driven Government

21

The Power of Setting Data Free

DATA.GOV
Citizen-Centered, Results Driven Government

H1N1 Widget

HHS 2009 Flu Info

DATASET SUMMARY

Agency:	Centers for Disease Control and Prevention
Category:	Health and Nutrition
Date Released:	
Date Updated:	
Time Period:	Daily (M–F)
Frequency:	Daily
Description:	This H1N1 widget provides a graphic with links to H1N1 Flu Info

DATASET RATINGS

	Current	Your Rating [?]
Overall	★★★★☆ (10 votes)	⊟ ★★★★★
Data Utility	★★★★☆ (9 votes)	⊟ ★★★★★
Usefulness	★★★★☆ (9 votes)	⊟ ★★★★★
Ease of Access	★★★★☆ (9 votes)	⊟ ★★★★★

deuce   serene

Type the two words:

DATASET INFORMATION

Data.gov Data Category Type	Tool Catalog
Specialized Data Category Designation	Widget/Gadget
Keywords	federal widget, H1N1, flu information, swine, origin, united states information, international information, things to do, plan, prepare, planning, preparing, virus, outbreak, Novel Influenza A, hospitalizations, deaths, unsubtypable
Unique ID	12

DOWNLOAD INFORMATION

Data Extraction ▶

Widget ▶   ⚙

Comment on these data:

SUBMIT

Click on Widget

Policy)

DATA.gov

Citizen-Centered, Results Driven Government

23

H1N1 Widget to CDC.gov

Get real time updates on your phone

Promote health awareness

Track the H1N1 flu and other health alerts

DATA.gov

Citizen-Centered, Results Driven Government

24

Prevent Spreading of Illnesses

FDA Peanut Product Recall Widget

FDA Product Recall List

FDA Salmonella
Peanut Product Recall

Information current as of
12 Jun 2009 16:00:00 GMT

Search [   ] [GO]

Brownie Product Recalls

Cake and Pie Product Recalls

Candy Product Recalls

Cereal Product Recalls

Cookie Product Recalls

Cracker Product Recalls

Donut Product Recalls

Dressing and Seasoning Product Recalls

The FDA Peanut-Containing Product Recall widget allows you to browse the Food and Drug Administration (FDA) database of peanut butter and peanut-containing products subject to recall. This database makes it easier for you to determine whether any of the products you have at home are subject to recent recalls, and will be updated as new information becomes available.

Copy This Code for FDA Peanut-Containing Product Recall:

```
<h3 id="start-widget-focus" style="margin-top: 10px; margin-bottom:-10px;"><a href="http://www.accessdata.fda.gov/scripts/peanut butterrecall/index.cfm" style="margin-left: 40px; font-weight: bold;" title="Jump directly to the
```

[+] Add to Google
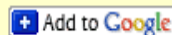
FDA Peanut-Containing Product Recall

DATASET SUMMARY

Agency:	US Food and Drug Administration
Category:	Health and Nutrition
Date Released:	
Date Updated:	
Time Period:	
Frequency:	
Description:	The FDA Peanut-Containing Product Recall widget allows you to browse the Food and Drug Administration (FDA) database of peanut butter and peanut-containing products subject to recall. This database makes it easier for you to determine whether any of the products you have at home are subject to recent recalls, and will be updated as new information becomes available.

DOWNLOAD INFORMATION

Data Extraction ▸

Widget ▸ ⚙

.gov

Exit

Comment on these data:

[   ]

SUBMIT

(Privacy Policy)

Click on Widget

Transparency: Track your Tax Dollars

USAspending.gov

	Contractors per page: 100	Page number: 1	Contractors 1-100 of Total 46981	First	Prev

Contracts from Dept. of Defense (FY 2009)

Parent Company Name	Contractor	
Lockheed Martin Corporation	ACS DEFENSE;ACS GOVERNMENT S SYS;ASPENMED SVC;COHERENT TECH;EAGLE GROUP INTL;GENERA...more	$16,186,056,381
The Boeing Company	;ALTEON TRAINING;ALTEON TRAINING 0001;AUTOMETRIC;AVIALL;AVIALL SVC;BOEING;BOEING 0674;BOEING 3;BOEIN...more	$13,876,918,725
Northrop Grumman Corporation	ADAPTIVE OPTICS ASSOCIATES;AEROJET ELECTRO SYS;AEROJET GENERAL;AMHERST SYS;AMSEC;AMSEC LIMITED LIABI...more	$7,848,633,795
	ANCED INFORMATION ENGINEERI;ADVANCED INFORMATION EERING SVC;AMERICAN OVERSEAS MARINE;AMERICA...more	$7,461,323,067
	STON ASSOCIATES;IDS RAYTHEON RF COMPONENTS;JPS UNICATIONS;OAKLEY NETWORKS;PHOTON RESEARCH .more	$5,650,315,592
	;AHLBRANDT TECHNISCHE ANLAGEN GMBH;AMTECH ATOR SVC;ANTRONNIX ANTENNA;ARCHITECTURAL GY;ALL more	$4,346,198,965

DATA.GOV

HOME | CATALOGS | STATE/LOCAL | ABOUT | FAQ | CONTACT US | SUGGEST OTHER DATASETS

USA Spending Contracts and Purchases

DATASET SUMMARY

Agency:	General Services Administration
Category:	Federal Government Finances and Employment
Date Released:	2008
Date Updated:	2009
Time Period:	2009
Frequency:	Monthly
Description:	Provides US Federal Spending by agency. The Contracts section allows searching for data relating to all Federal contracts. There are three main categories of searches: by contractor, by location, and by contracting agency. The Purchase section allows users to view purchase cards spending of various federal agencies for each fiscal year. The data available on this website are extracted from the rep (more)

DOWNLOAD INFORMATION

Data Extraction ▸

Widget ▸

Comment on these data:

SUBMIT

(Privacy Policy)

Click on Data Extraction

Save Time During Air Travel

Legend
- General Arrival/Departure delays are 15 minutes or less.
- Departures are experiencing taxi delays of 16 to 45 minutes and/or arrivals are experiencing airborne holding delays of 16 to 45 minutes.
- Traffic destined to this airport is being delayed at its departure point. Check your departure airport to see if your flight may be affected.
- Departures are experiencing taxi delays greater than 45 minutes and/or arrivals are experiencing airborne holding delays greater than 45 minutes.
- This denotes a closed airport!

View by Region: Select a Region ▶

Search by Airport: Airport Lookup ▶
(Enter city, airport code, airport name)

View by Major Airport: Select a Major Airport ▶

THURSDAY, JUNE 25, 2009   Text A⁺ A

DATA.GOV

HOME | CATALOGS | STATE/LOCAL | ABOUT | FAQ | CONTACT US | SUGGEST OTHER DATASETS

Airline On-Time Performance and Causes of Flight Delays

DATASET SUMMARY	
Agency:	Bureau of Transportation Statistics
Category:	Transportation
Date Released:	10/1/1987
Date Updated:	5/12/2009
Time Period:	October 1987 onward
Frequency:	monthly
Description:	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 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.

DOWNLOAD INFORMATION
Data Extraction ▶
Widget ▶

Comment on these data:

SUBMIT

(Privacy Policy)

Click here to set data parameters

DATA.GOV
Citizen-Centered, Results Driven Government

27
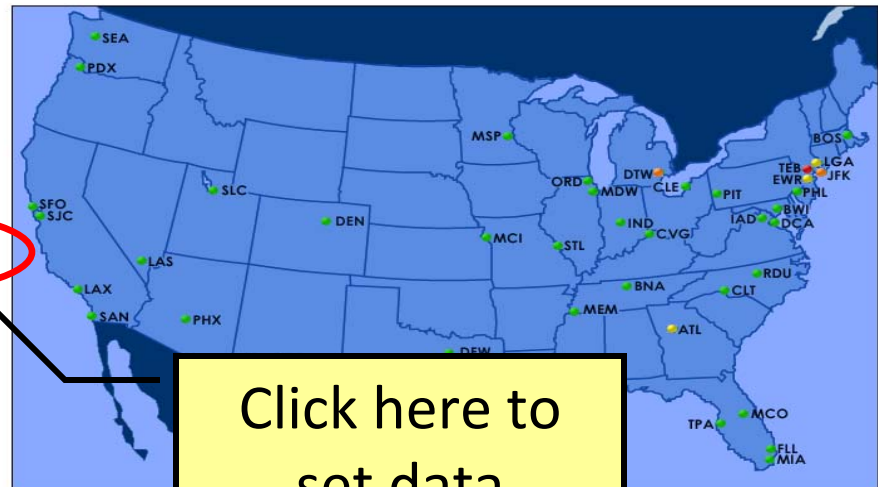
......Get there on time

Using data from the Federal Government, we can tell you how late your flight is on average or help you find the most on-time flight from one city to another.

FlyOnTime.us

Airline:

From: (city or airport)

Flight #: (optional)

OR

To: (city or airport)

Search >>

Search >>

Terms of Use

DATA.GOV
Citizen-Centered, Results Driven Government

Aviation Data & Statistics

Federal Aviation Administration

Air Traffic | **Data & Research** | Training & Testing | Licenses & Certificates | Regula...

FAA Home » Data & Research » Aviation Data & Statistics

Aviation Data & Statistics

Updated: 12:04 pm ET March 24, 2009

Aeronautical
- Aeronautical Survey Program

Airlines & Airports
- Air Operators Utilization Reports
- Airline On-Time Statistics & Delay Causes
- Airport Data (5010) and Contact Information
- Airport Noise Compatibility Planning Data and Program Status
- Airport Obstruction Charts (NOAA)
- DOT Airline Statistics
- FAA Operations & Performance Data
- Forecasting Aviation Activity by Airport (MS Word)

Airmen & Aircraft
- Airmen Knowledge Test Statistics
- Data Downloads
- Service Difficulty Reporting
- U.S. Civil Airmen Statistics

General Aviation
- General Aviation and Air Taxi Activity Surveys
- Model for Estimating General Aviation Operations at Non-Towered Airports (MS Word)

Other Resources
- Data Library (DOT)
- System Standards

DATA.gov
Citizen-Centered, Results Driven Government

Summary Statistics
Origin Airport

BUREAU OF TRANSPORTATION STATISTICS

Origin Airport: Washington, DC - Washington Dulles International (IAD)
Airline: Delta Air Lines (DL)
Time Period: January 1, 2009 to January 1, 2009

Note: A complete listing of airline and airport abbreviations is available. Times are reported in local time using a 24 hour clock.

Excel | CSV

Carriers	All Flights				Late Flights				Total Number Cancelled	Percent Flights Cancelled	Total Number Diverted	Percent Flights Diverted	Percent Flights Late
	Total Number	Average Departure Delay (minutes)	Average Taxi-Out (minutes)	Average Scheduled Departure to Take-off (minutes)	Total Number	Average Departure Delay (minutes)	Average Taxi-Out (minutes)	Average Scheduled Departure to Take-off (minutes)					
ALL	8	16.12	15.43	31.55	40	82.22	15.25	97.48	0	0.00	0	0.00	21.98
DL		-5.86	12.86	7.00	0	0.00	0.00	0.00	0	0.00	0	0.00	0.00
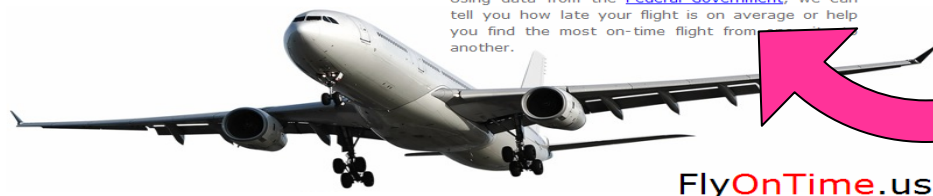
Using data from the Federal Government, we can tell you how late your flight is on average or help you find the most on-time flight from one airport to another.

FlyOnTime.us

From: [   ] To: LAX Day: [   ] Search >>

Most On-Time Flights and Airlines
Data from April, 2009

From **Washington, DC (IAD)** to **Los Angeles, CA (LAX)**:

Flight	Averge Arrival	Num Flights
United 925	10.9 min. early	30
United 49	7.8 min. early	30
United 83	3.9 min. early	30
United 209	3.8 min. early	30
American 75	3.8 min. early	29
United 947	2.5 min. early	29
American 149	2.5 min. early	30
United 867	1.4 min. early	30
United 841	4 min. late	26
United 967	4.8 min. late	29
American 263	9.6 min. late	30

Airline: [   ]

Flight #: (optional) [   ]

OR

From: (city or airport) [   ]

To: (city or airport) [   ]

Search >>   Search >>

Terms of Use

"A new reframing of the world wide web"

Innovative Problem Solving

Open government data

Tap the Ingenuity of the American People

DATA.GOV
Citizen-Centered, Results Driven Government

30
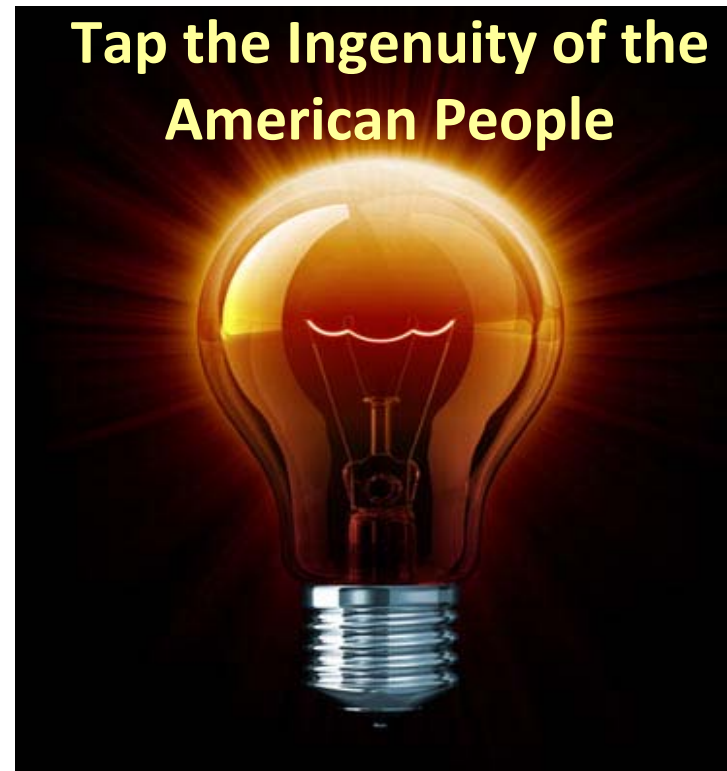
Engage and Participate

- Blog

- Applications

- Projects

Comments

Markdown syntax is supported in comments.

Robert Kosara

Works with the command-line unzip utility on the Mac too, so it should work in Linux as well. These .exe files are usually just ZIP files with a little expander program tacked on; they're not as bad as they seem. Yeah, I'm the apologist again, sorry ;) But this is actual data, not PDFs or anything.

Robert Kosara

The much bigger problem is that most of the data inside these ZIPs/EXEs is .doc and .xls files, rather than .pdf and .csv.

Michael E. Driscoll

I think this points to the difference between making data available versus making it accessible.

Making raw data available isn't hard -- you simply upload whatever bundle of bits (in this case, a bunch of .exe files) and voila, you're finished.

But making data accessible has a tried and true name: publishing. Publishing is hard, and publishing data is arguably harder, as anyone who has ever prepared a data set to share with colleagues, collaborators, or just to post on a web site can tell you.

Data.gov is a good idea, and I'm a big fan of the initiative, but it will have to overcome some of these burdens if it's going to gain traction.

Pete Skomoroch

I'm going to side with Data.gov on this one.

The last thing we should do is discourage them from putting more data up quickly... just label it as "other" instead of CSV. If someone writes a parser, they should

Citizen-Centered, Results Driven Government

Apps for AMERICA 2
the data.gov challenge

Google

O'REILLY®

Craig Newmark

gov2.0 EXPO SHOWCASE

techweb™
A Division of United Business Media LLC

RECENTLY SUBMITTED APPS

Top Dangerous Mines
A website showing a real-time list of the most dangerous mines in the US based on recent earthquake data.
Submitted June 18, 2009 by Benjamin Hardin

FlyOnTime.us
Find the most on-time flight from one city to another.
Submitted June 11, 2009 by Josh Sulkin

FBI Fugitive Concentration
Play Concentration with photos of FBI Fugitives
Submitted May 22, 2009 by Bertrand Fan

DATA.GOV
Citizen-Centered, Results Driven Government

Agency Participation and Citizen Engagement are the Keys to the Success of Data.gov

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

Citizens Using Data.gov Data to Create Applications and Mashups

Citizens Suggesting Datasets They Want To See

Citizens Commenting on Specific Datasets

Data.gov Users Rating Datasets Using a Five Star Rating

DATA.GOV
Citizen-Centered, Results Driven Government

The Future of Data.gov….

- 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.

DATA.GOV
Citizen-Centered, Results Driven Government

Questions?

The days of searching long and hard for data are over…

..Data is now brought to you and more easily available..

…And should not be such a "bear" of a problem!

DATA.GOV
Citizen-Centered, Results Driven Government

35