Knowledge and Data Sharing

Presented to:
American Library Association
2010 Annual Conference

Presented by Joey Hutcherson, PMP, TPM,
Office of Chief Information Officer, U.S. Department of Commerce
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
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 is part of the Open Government Initiative from the Executive Office of the President
• Provides a way for government organizations to share data on performance and their mission with the public
• Seeks to drive innovative uses of these datasets to help people every day
• Creates a knowledge management capability across the government (central catalog)
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
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!
When the Department of Defense released its satellite data... the public created affordable GPS devices!
• 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
Goals of Data.gov

• 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
• 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
What is Meant by Raw Data?

- 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
Welcome to Data.gov
Applications

• 237 applications developed by citizens and others to create value from the raw data provided
Using Dept. of Transportation data, FlyOnTime.us lets people find the most on time flights and shortest security line wait times.

Integrated to Twitter.
National Obesity: Eat Your Vegetables

- Uses Health and Human Services community health data and shows obesity rates by county
- Supports the First Lady’s fights against childhood obesity

http://public.tableausoftware.com/views/contributorstoobesity/Eatyourvegetables?embed=yes&:toolbar=yes
Fix My City DC

- Local government is helping as well, The District of Columbia’s Open311 API powers FixMyCityDC.com
- Lets people report problems and see what’s been reported and responded to throughout the District
• Using Bureau of Labor local area unemployment data
• This app lets people explore which cities and neighborhoods have the best growth in new jobs

http://pujaplicaciones.javeriana.edu.co/Employment/
• Using data from the Environmental Protection Agency, this app lets you search by zip code or Google map interface to find environmental issues in your neighborhood
• Links directly to detailed EPA pages for more information and how to report an issue
• [http://checkitnsee.com/](http://checkitnsee.com/)
Data Masher

- Uses data from many agencies
- Allows anyone to create a mashup between two datasets to see potential correlations
- [http://www.datamasher.org/](http://www.datamasher.org/)
Semantic Technologies

- Providers developers
  the tools and raw data
  formats they need to
  develop new capabilities

- Close partnership with
  Rensselaer Polytechnic
  Institute

- Connected to other open
  data efforts across the
  world
Catalogs

Data.gov Catalogs

Use the Data.gov catalog below to access U.S. Federal Executive Branch and agency datasets. By accessing the catalog, you can search for datasets by keyword, browse the catalogs, or rate the tool. Each catalog provides an instant download of machine readable, platform-independent datasets, while the Tool Catalog allows you to mine datasets.

Search tools by keywords

Search tools by single/multiple category

Search tools by single/multiple agency

Name (click for metadata and to rate tool)

Rating Agency Category Data Extraction Widget

Airline On-Time Performance and Causes of Flight Delays
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...

American FactFinder
American FactFinder is the U.S. Census Bureau's online, self-service tool designed to search the American Community Survey, Puerto Rico Community Survey, Demographic Census, ...

Cancer Incidence - Surveillance, Epidemiology, and End Results (SEER) Registries Limited-Use
SEER Limited-Use cancer incidence data with associated population data. Geographic areas available are county and SEER registry. The Surveillance, Epidemiology, and End ...

CDC Emergency Text Messages
Provides up-to-date information on public health text alerts before and during emergencies.

Dataferret
The Dataferret tool is used to search American Community Survey Public Use Microdata. Current Population Survey (CPS); CPS Supplemental Consumer Survey of Income and Pro...

Input keywords

Rate the tool
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:** 8/19/2009
- **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 managed by certain industries as well as federal facilities.

**DATASET RATINGS**
- **Overall**: 
  - Current: 4 stars
  - Your Rating: 5 stars (2 stars)
- **Data Utility**: 
  - Current: 4 stars
  - Your Rating: 5 stars (3 stars)
- **Usefulness**: 
  - Current: 4 stars
  - Your Rating: 5 stars (3 stars)
- **Ease of Access**: 
  - Current: 4 stars
  - Your Rating: 5 stars (3 stars)

**DATASET INFORMATION**
- **Data.gov Data Category Type**: Raw Data
- **Specialized Data Category Designation**: 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

**DOWNLOAD INFORMATION**
- **Format Options**: CSV/TXT, XML, XLS, Shapefile
- **File Size**: 37.4 MB

- **Download data here**
- **Rate data here**
- **Comment here**
Key Enabler for “Killer Apps”

- “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
Global Community

- Made of
  - Communities of Practice
  - Communities of Interest
  - Educators and students
  - Developers
  - Local, state, and tribal governments
  - Other governments
Top Ten States Visiting Data.Gov

- California
- District of Columbia
- New York
- Texas
- Virginia
- Illinois
- Massachusetts
- Washington
- Florida
- Georgia
Top Ten Countries Visiting Data.gov

- USA
- Canada
- China
- United Kingdom
- Germany
- Australia
- France
- Netherlands
- Italy
- India
“A new reframing of the world wide web”

Innovative Problem Solving

Tap the Ingenuity of the American People

Open government data
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 website 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 be able to do it. I think better data is always welcome.
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 Sultin

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

| DATASET RATINGS | Current | Your Rating [?]
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>Data Utility</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>Usefulness</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
<tr>
<td>Ease of Access</td>
<td>🌟🌟🌟🌟🌟</td>
<td>🌟🌟🌟🌟🌟</td>
</tr>
</tbody>
</table>
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, mashups, and applications from the data on Data.gov
- Linked open data...semantic web, etc.
What’s Next

• Year one focused on building the intra- and inter-government community
• Year two will focus on building communities with
  – Developers
  – Educators
  – Students (K-12 and universities)
  – Private industry
  – International
• The open government and linked data movement has been ignited by Data.gov
• 7 countries have versions of open data platforms and more are coming
Empowering People

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

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

...To let people make more informed decisions every day!
Questions

Contact Information:
Joey A. Hutcherson, PMP, TPM
Information Technology Infrastructure Management
Office of the Chief Information Officers,
Office of the Secretary
U.S. Department of Commerce
Email: jhutcherson@doc.gov