I Hope the Public Finds Problems with Government Open Data Because That Is the Point

Association of Public Data Users Conference
Tuesday September 12, 2016 | Alexandria, VA
Rebecca Williams, GovEx
Rebecca Williams

@internetrebecca
rebecca.williams@jhu.edu
govex.jhu.edu
whatworkscities.bloomberg.org
1. Not Garbage Going In (Validation)
DATA Act Broker - Beta

Welcome to the DATA Act Broker – Beta version.

Sign in to upload your agency financial data and validate it against the DATA Act Information Model Schema (DAIMS) v1.0.

Visit the Data Submission page on MAX for details on how to format your data.

https://alpha-broker.usaspending.gov
Validator

There are three ways you can validate data.json, either by validating a public URL, uploading a json file, or pasting the raw JSON into the form.

Validate data.json URL

**Schema** [Federal v1.1](#)

- View in Browser
- Output JSON

**data.json URL**

[Example: http://energy.gov/data.json](http://energy.gov/data.json)

[Validate URL](#)

Validate data.json file upload

**Schema** [Federal v1.1](#)

**Upload a data.json file**

- Choose File
- No file chosen

[Validate File](#)

Validate raw JSON

[https://labs.data.gov/dashboard/validate](https://labs.data.gov/dashboard/validate)
2. Show Your Work (Documentation)
Schedule E covers the line item expenditures for independent expenditures. For example, if a super PAC bought ads on TV to oppose a federal candidate, each ad purchase would be recorded here with the expenditure amount, name and id of the candidate, and whether the ad supported or opposed the candidate.

An independent expenditure is an expenditure for a communication "expressly advocating the election or defeat of a clearly identified candidate that is not made in cooperation, consultation, or concert with, or at the request or suggestion of, a candidate, a candidate's authorized committee, or their agents, or a political party or its agents.

Aggregates by candidate do not include 24 and 48 hour reports. This ensures we don't double count expenditures and the totals are more accurate. You can still find the information from 24 and 48 hour reports in /schedule/schedule_e/.

Operations

+ /committee/{committee_id}/schedules/schedule_e/by_candidate/
+ /schedules/schedule_e/
+ /schedules/schedule_e/by_candidate/

https://api.open.fec.gov/developers/

E-filing endpoints provide real-time campaign finance data for electronic filers.

These endpoints are perfect for watching filings roll in when you want to know the latest information. E-filing endpoints only contain the most recent data.
We're designing this site in phases. This allows us to launch new features faster than releasing a completely redesigned FEC.gov. The official site for FEC data is still the Campaign Finance Disclosure Portal.

New features

https://beta.fec.gov/glossary
3. Teachable Moments (Tutorials & Training)
With tens of thousands of datasets ranging from satellite imagery to material standards to demographic surveys, the U.S. Department of Commerce has long been in the business of Open Data. Through the Commerce Data Usability Project, go on a series of guided tours through the Commerce data lake and learn how you can leverage this free and open data to unlock the possible.

https://www.commerce.gov/datausability/
Data Academy

A training program for City and County of San Francisco staff

Welcome to the City and County of San Francisco's Data Academy! Our goal is to help analysts citywide explore, refine, and enhance their skills in data analysis and visualization. We also offer classes for those hoping to increase their comfort working with data. We offer a series of classroom based courses detailed below and have also curated a list of courses offered by other institutions for free or at cost.

Course Catalog

<table>
<thead>
<tr>
<th>Topics</th>
<th>Introduction to R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This introductory class, participants will learn practical skills necessary for reading data into R from external sources, manipulating the data, and visualizing it. The course will also cover other basic programming concepts as they are implemented in R.</td>
</tr>
</tbody>
</table>

The Data Academy is produced in partnership with the Controller's Office.

Leveraging the City's Open Data Portal

This course helps you unlock features of the open data portal to improve your own data workflows. To get the most out of this class, you should have Excel 2010 with the Power Query Add-in (free) or Excel 2013 on your domain.

https://datasf.org/academy/
3. Vote on Data (UX Testing & Policy)
responses that had only a single user select them. Single responses are worth monitoring but not yet a trend.

**Task 1:**

You want to learn more about what types of reports candidates need to file and when. Where would you go?

This result is from our first test, where the most common incorrect paths were through Campaign finance data and Education and training. In our second iteration, we removed Education and training and added a Library of publications instead. We’ve also added more clear paths on the beta site between the Campaign finance data section and the Registration and reporting section.

**Task 2:**

https://github.com/18F/openFEC/issues/1809
Instead of guessing, we decided to go out and ask them.

Following our practice of focusing on user needs, the College Scorecard team first spent time engaging directly with users at every single step in the project. Some of our favorite conversations were when we talked to high school students in Washington, D.C.’s Anacostia neighborhood and their excellent guidance counselor. He told us how he “hacks” the process to make sure every one of his students gets an acceptance letter from a college, and feels that rush of possibility.

We met with 38 4-H kids from across the country who were frustrated by https://collegescorecard.ed.gov/ that she shopped for schools based on the lowest annual actual cost to
Be a tester, get paid

The Civic User Testing Group (CUTGroup) is a community of residents of Chicago and all of Cook County who get paid to test out civic websites and apps. This is how it works:

- Fill out the CUTGroup profile form below to sign up to be a tester. Once you do that, we will send you a $5 VISA gift card just for signing up.
- Once you sign up, you will receive email notifications about new test opportunities. We will ask a few more questions that relate to the

http://www.cutgroup.org/
PROJECT OPEN DATA

Open Data Policy – Managing information as an Asset

1. Background
Data is a valuable national resource and a strategic asset to the U.S. Government, its partners, and the public. Managing this data as an asset and making it available, discoverable, and usable – in a word, open – not only strengthens our democracy and promotes efficiency and effectiveness in government, but also has the potential to create economic opportunity and improve “quality” of life.

For example, when the U.S. Government released weather and GPS data to the public, it fueled an industry that today is valued at tens of billions of dollars per year. Now, weather and mapping tools are ubiquitous and help everyday Americans navigate their lives.

The ultimate value of data can often not be predicted. That’s why the U.S. Government released a policy that instructs agencies to manage their data, and information more generally, as an asset from the start and, wherever possible, release it to the public in a way that makes it open, discoverable, and usable.

The White House developed Project Open Data – this collection of code, tools, and case studies – to help agencies adopt the Open Data Policy and unlock the potential of government data. Project Open Data will evolve over time as a community resource to facilitate broader adoption of open data practices in government. Anyone – government employees, contractors, developers, the general public – can view and contribute. Learn more about Project Open Data Governance and dive right in and help to build a better world through the power of open data.

2. Definitions
This section is a list of definitions and principles used to guide the project.
2-1 Open Data Principles - The set of open data principles.
2-2 Standards, Specifications, and Formats - Standards, specifications, and formats supporting open data objectives.
2-3 Open Data Glossary - The glossary of open data terms.
2-4 Project Open Data Metadata Schema - The schema used to describe datasets, APIs, and published data at agency.gov/data.

3. Implementation Guidance
Implementation guidance for open data practices.
these elements will be standardized across the different types of awards (grants, contracts, etc.).

<table>
<thead>
<tr>
<th>Data Element Name</th>
<th>Legislation</th>
<th>Provide Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awarded/Recipient Legal Business Name</td>
<td>FFATA</td>
<td>feedback period closed</td>
</tr>
<tr>
<td>Awarded/Recipient Unique Identifier</td>
<td>FFATA</td>
<td>feedback period closed</td>
</tr>
<tr>
<td>Ultimate Parent Unique Identifier</td>
<td>FFATA</td>
<td>feedback period closed</td>
</tr>
<tr>
<td>Ultimate Parent Legal Business Name</td>
<td>FFATA</td>
<td>feedback period closed</td>
</tr>
<tr>
<td>Legal Business Address</td>
<td>FFATA</td>
<td>Leave feedback (closes 2/29/15)</td>
</tr>
<tr>
<td>Legal Business Congressional District</td>
<td>FFATA</td>
<td>Leave feedback (closes 2/29/15)</td>
</tr>
<tr>
<td>Legal Business Country Code</td>
<td>FFATA</td>
<td>Leave feedback (closes 2/29/15)</td>
</tr>
<tr>
<td>Legal Business Country Name</td>
<td>FFATA</td>
<td>Leave feedback (closes 2/29/15)</td>
</tr>
<tr>
<td>Top 5 Highly Compensated Officer Names</td>
<td>FFATA</td>
<td>feedback period closed</td>
</tr>
<tr>
<td>Top 5 Highly Compensated Officer Total Compensations</td>
<td>FFATA</td>
<td>feedback period closed</td>
</tr>
</tbody>
</table>

**Award Amount Information**

This list includes elements used to describe the amount (value) of a certain award.

<table>
<thead>
<tr>
<th>Data Element Name</th>
<th>Legislation</th>
<th>Provide Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding Action Obligation</td>
<td>FFATA</td>
<td>Leave feedback</td>
</tr>
</tbody>
</table>
Policy fights of the future will be about fields and methodologies.
5. It Just Works! (Plug-and-playable data)
Draft Geospatial Open Data Standards

Last November, Mayor de Blasio signed Local Law 108 of 2015, mandating the formation of a working group tasked with creating standards for address and geospatial information on the Open Data Portal. Over the last several months, the working group drafted recommendations for geospatial attributes, column headers, and geocoding processes for datasets on the Portal. In the spirit of Open Data for All, we invite the public to join the conversation as we finalize these recommendations into formal requirements.

Open Data promotes transparency by publishing data used internally in the City with as few transformations as possible. While we seek to maximize the usability of data on the Portal, the Open Data team changes the form of raw data only when it presents minimal risk of misrepresenting source information. As a result, the recommended geospatial standards reflect the data fields most frequently captured by City agencies, information that is in highest demand from public users, and attributes that will have the biggest impact on citywide operations once they are standardized.

Please read the standards and leave your comments, suggestions, and questions in the form below. Public comments will close on Thursday, September 15. The working group will consider all comments in finalizing the standards. Comments can be submitted via email to ryan.gilmore@nyccio.nyc.gov.
Reach out to Public Data Maintainers!

labs.data.gov/dashboard/
project-open-data.cio.gov/points-of-contact/
data.gov/contact
Rebecca Williams

@internetrebecca
rebecca.williams@jhu.edu