
 This repository Search

ExploreGistBlogHelp

dazzaji + Settings

Watch 18

Star 6


Fork 249

 **fedspendingtransparency** / **fedspendingtransparency.github.io**
forked from [project-open-data/project-open-data.github.io](#)

What Federal spending data elements are most crucial to your current reporting and/or analysis? #6

New issue

 **rmaziarz** opened this issue on Dec 1, 2014 · 10 comments




rmaziarz commented on Dec 1, 2014

Owner+1

No description provided.

 **rmaziarz** added the **data elements question** label on Dec 1, 2014




GaryBass commented on Dec 22, 2014

+1

I'm new to GItHub, but have been involved in spending transparency issues for many years. For example, when I ran OMB Watch, we developed FedSpending.org, which became the underlying programming for USAspending.gov. As noted below, there are many data elements and fields not covered by the 49 listed from FFATA. Here are some points about what is listed and what is not:

1. In Data Element #3 the word "Ultimate Awardee" is used. It's the only time the word "ultimate" appears in the 59 data elements. What does this mean? It would be helpful if subrecipient reporting did require the ultimate recipient to report, not simply one or two-tiers below the prime recipient. Thus, I would recommend developing data element standards for the ultimate recipient.
2. It is essential to have a parent company identifier. The DUNS standard is not public, is not very accurate, and isn't flexible enough for tracking federal spending. For example, if a parent company sells off a unit of the company, it is important to trace back earlier years to its previous parent company while still being able to track its new independence. The DUNS system does not allow for this type of historical analysis.
3. Under Data Element #49, what fields are envisioned for "business type"? It would be very useful to have type of for-profit (e.g., sole proprietor, S Corp) and nonprofit (e.g., 501(c)(3), 501(c)(6)). And to require the recipient to identify the type of business it is.
4. Why are the rest of the FFATA data elements not listed? Aren't standards needed for these fields? As a user, I would think so. Or is there an assumption that the other data elements in FFATA already have standards?
5. There are some data elements missing. For example, it would be very useful to have a CEO-median worker pay ratio. The Recovery Act had disclosure of CEO salary, but a better approach (as mandated under Dodd-Frank) is a CEO-worker pay ratio. It could allow companies to calculate the median worker pay using statistical sampling in order to reduce any potential burdens.



JimHarperDC commented on Dec 23, 2014

+1

Every data element is special, of course, but the bulk of use cases would probably benefit from giving priority to data elements that provide a "skeletal" view of the spending cycle as illustrated by Commissioner Lebryk in his [recent PowerPoint presentation](#).

What's below may or may not directly correspond with the list of federal spending data elements [proposed for discussion](#), but they are probably at least implicit requirements of the DATA Act. Its direct

Labels

Milestone

Assignee

Notifications

7 participants

Export Issue

requirements can't be fulfilled without them.

The discussion below is adapted (and possibly maladapted) from the appendices of the [Grading the Government's Data Publication Practices](#) study.

Budget Authority

The first data element would be **budget authority**. Budget authority is authority to obligate funds. (I'll discuss only appropriations - authorizations of appropriations are another form of budget authority that seems not to be relevant here. And this is budget authority in the "legal power" sense as opposed to the "amounts we have to spend" sense.) Such an authority has a unique identifier, which may be its location in the corpus of public laws, but whatever the case must uniquely identify each instance of authority to obligate funds. It has a fiscal year or years (or other time-period) during which it exists (or it has a no-year existence). It has an amount. It has a purpose, which is a string of legislative text that states its purpose. It may have an organizational unit of government that Congress has assigned to exercise the authority.

Organizational Units

What we classed as "ExecutiveAgents" in the "Grading" study appear to be "Funding Agency", "Funding Agency Sub Tier" or "Awarding Agency" and similar data elements in the list. In our effort, we used the organizing terms: "Agency", "Bureau", "Program", and "Project", but the concepts can be applied to any hierarchy of organizational units.

A top-most organizational unit will have, at a minimum, a "name," which is a string of characters, such as "Department of the Treasury," and a unique identifier (such as the OMB agency code). A second-tier organizational unit will have a "name" and a unique identifier, which is best formulated if it implies the top-level organization of which it is a part, such as by incorporating the unique identifier of the parent. A third-tier organizational unit will have a "name" and unique identifier implying it's parent and grandparent. Lower organizational units will similarly have names and unique identifiers that imply their place in the organization of the executive branch.

Obligation

An **obligation** is a binding agreement or statutory requirement that will result in outlays. An obligation should have a name, which is a human-readable string of characters, and a unique identifier. It may have a summary, and it may have a document or text which captures all or most terms of the obligation (i.e. a copy of the grant or contract). It has an organizational unit, which is the entity entering into or subject to the obligation and responsible for extinguishing or carrying it out via outlay. And it has a budget authority, which is the authority under which the obligation is entered into. An obligation may have one of several types, including: award, procurement contract, grant, salary, direct payment, and so on.

Party

A **party** is the recipient of an Outlay or non-federal party to an Obligation. This may include federal entities, state entities, federal employers, contractors, grant recipients, or foreign countries. Represented by a number of entities put forward in the discussion list, parties have at least names and unique IDs.

Outlay

An **outlay** is spending in execution of an Obligation. It has a unique identifier (a transaction ID, like the unique identifiers used in the Treasury Transaction Reporting System). It has a budget authority. It has an obligation. It has an amount (which in rare cases can be negative, signifying a credit to the relevant treasury account). It has a Treasury account ID, and Treasury Sub-Account Id. It has an Organizational Unit, the "holder" of the Treasury Account/Sub-Account using the outlay to fulfill/extinguish the obligation. It has a transaction date and a settlement date. And it has a party, the recipient of the outlay (which can be omitted for privacy reasons when the payee is an individual, private recipient of benefits, for example).

Standardizing and requiring use of these as minimum elements would probably produce the most benefit to the most data users. It is bare-bones and would completely satisfy almost nobody. The central challenge in the current effort may be to maintain focus, though, and not to indulge in the temptation to perfect any data element or set of elements for any use case. There are always more data elements, and more properties you can add to existing data elements, until you have tasked yourself with describing the entire universe. Describing the bare bones of the spending cycle is a way of sanely scoping the first phase of the effort. Later phases can add further elements and detail existing elements.

Have fun!



Solomon commented on Dec 23, 2014

+1

I'm not as close to the data as **@GaryBass** and **@JimHarperDC** seem to be, but I'll tell you my use case, and hopefully that can help guide your thinking on some of the data elements that would be necessary.

Every year we publish the federal budget with data about how the US spends and makes money. That budget has about 4,000 line items, and has a hierarchy of Departments, Bureaus, and Agencies.

An example of a line item in this context would be in 2013, within the Department of Education, in the bureau of Office of Innovation and Improvement, we spent \$4.3 billion on "Innovation and Instructional Teams".

I built an interactive visualization to explore the US Budget, which you can find here: http://solomonkahn.com/us_budget/. Unfortunately, there is no way for me to link that \$4.3 billion in rolled up budget spending on "Innovation and Instructional Teams" to the specific spending that combined to make up the total \$4.3 billion in that line item. (or if there is, I would be ecstatic to learn how this would be possible)

I would like to extend this visualization so that people can see what specific spending makes up the line items in the US Budget. As you create data standards for federal spending, I would like you to take into account how these different sources could be connected, and how I could take the data from federal spending and link it to the US Budget.

Feel free to contact me anytime to discuss this further.



nsinai commented on Dec 23, 2014

+1

Great feedback **@GaryBass** **@JimHarperDC** **@Solomon**



RehanaMohammed commented on Dec 24, 2014

+1

@Solomon thanks for your comment. We are exploring ways to make budget and spending data more accessible and useful, including what you describe.



RehanaMohammed commented on Dec 24, 2014

+1

@JimHarperDC thanks for your comment. We will take a look at that presentation you posted. Agree with your last point - we are adopting a similar approach and focusing on integral elements first and working in an iterative way, which is reflected in the list of elements.



RehanaMohammed commented on Dec 24, 2014

+1

@GaryBass thanks for those questions. We'll post some responses in the new year.



bsweger commented 14 days ago

+1

Just want to second **@Solomon**'s great input and add my own .02.

During my time at [National Priorities Project](#), an org that makes the federal budget accessible to regular citizens, I spoke with many partners and users about federal spending data, and they mostly want to drill into it in two ways:

- What the money's being used for (as specific as possible, per **@Solomon**'s point)
- Geography (country, state, county, congressional district, zip code)

Many people who need this information do issue-area advocacy and don't have resources dedicated to budget analysis and data wrangling. They don't differentiate between "programs," "projects," or other

words that many of us struggle to define. They simply want to answer questions like “How much Head Start money did my state/city get last year? Is that more or less than last year?”

As you know, the Head Start question can be answered via some federal spending data (USASpending, where you can get it via the CFDA program number) but not in other datasets like outlays and budget authority, which roll Head Start into a larger Children and Family Services Programs account and don't provide geographic information.

Thus, even when we can get the details people want, we can't put those details into the context of the overall federal spending lifecycle.

Of course this disconnect isn't news to Treasury/OMB, but I just wanted to say that there's a demand for the work you're doing. The DATA Act isn't on the radar of many issue area people who will ultimately stand to benefit, but this constituency is just as important as the data vendors who hope to gain from standardized spending data.

I have some more data element-specific questions/comments but will post them separately.

Thanks!



bsweger commented 12 days ago

+1

The advice of **@JimHarperDC** to “sanely scope the first phase” makes a lot of sense. His proposed “skeletal view,” if implemented correctly, would help people like **@Solomon** drill into specific line items (e.g., programs/projects/organizational units).

I would add geographic information about outlay recipients to his list of data elements that will provide the most benefit to the most data users (happy to see the address fields on your data elements list).

Lastly, I have a question about federal loans and federal insurance. It's always vexing to figure out how much money is really spent on those programs (versus the contingent liability amounts that we currently see in USASpending). In these cases, would the outlays field represent actual spending?



nsinai commented 4 days ago

+1

Great comments **@bsweger**



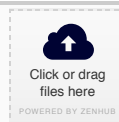
Write

Preview

Markdown supported

Edit in fullscreen

Leave a comment



Attach images by dragging & dropping, [selecting them](#), or pasting from the clipboard.

Comment

