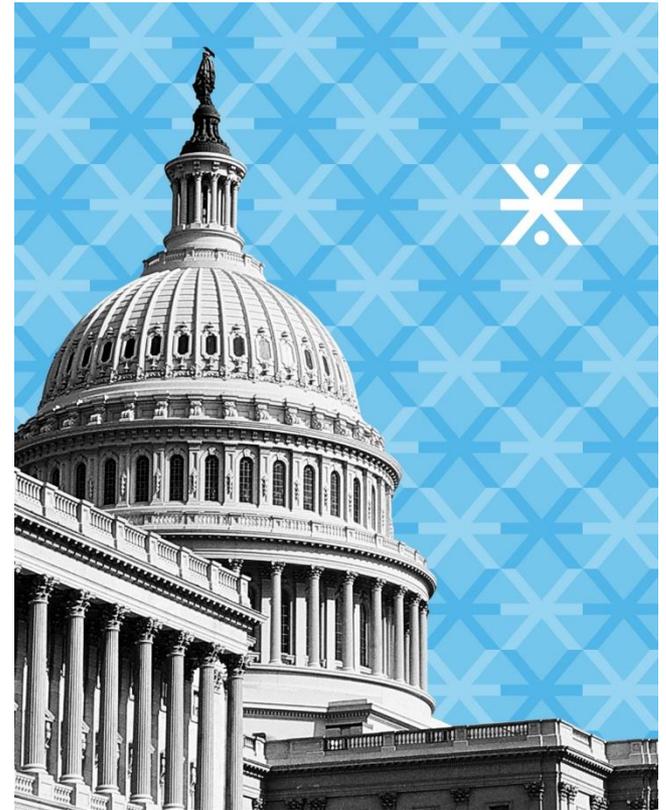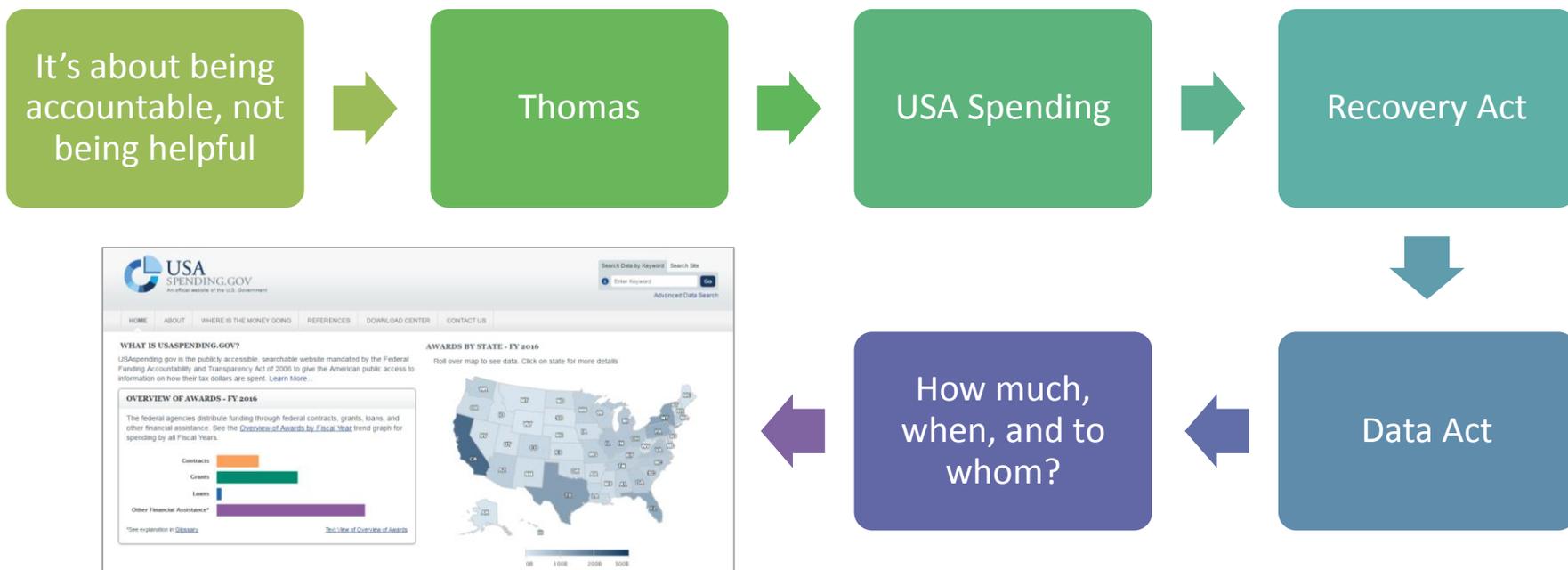# Summit Consulting

Building Demand for Open Data:

A Federal Perspective

September 13, 2016

What Does Data Transparency Mean to a Federal Policy Maker?

Unlike many state and municipal governments, Federal agencies are often obligated by *data transparency* mandates to release their data.

It's about being accountable, not being helpful → Thomas → USA Spending → Recovery Act → Data Act → How much, when, and to whom?

Consider expanding the Federal perspective on data transparency beyond meeting good government requirements to include both:

– Supporting the agency's mission

– Not risking the agency's reputation

Releasing Data to Promote Mission

Public data release can support the agency's mission in several ways:

Facilitate secondary analyses

Facilitate secondary analyses that can provide insights to improve the agency's programs and policies.

Improve quality

Allow feedback from data users that can help improve data quality.

Become an authoritative resource

Establish your organization as an authoritative resource for researchers.

Realize mission

Some agencies and regulators already use data as a _direct_ aspect of their mission.

Questions to Ask before You Release Administrative Data

To ensure that your public data release supports the agency's mission, you should ask the following questions:

Why	Who	What
Why are you publicly releasing the data?	Who is the audience for this data?	What kinds of data will you release?

Why Are You Publicly Releasing the Data?

- **What are the benefits of releasing the data both for the data-using public and for your organization?**
  - Example public benefits: Supporting increased civic engagement; facilitating general research
  - Example organization benefits: Encouraging user feedback to improve your data collection efforts, program performance, or program impact
- **Understanding the reasons for releasing the data will help determine the audience for the data and the kinds of data to release.**
  - For instance, if your reason is to satisfy accountability mandates, your audiences are likely the media and general public and the kinds of data you release mostly focus on cost and performance information.

Who Is the Audience for the Data?

Media & General Public

- Narrower interests, focusing on information related to government accountability
- Fairly limited data skills and experience

Policymakers & Practitioners

- Interests focus on program cost and performance, but want more detailed information
- More skill with data analysis and statistics

Researchers & Scholars

- Widest range of interests
- Most skilled users of data

What Are Administrative Data?

- **Collected in the course of regular recordkeeping, typically tracking participants, registrants, employers, or transactions—not collected for research purposes**

- **Rich with information that is frequently updated and requires no additional expense for data collection (often referred to as "lazy data")**
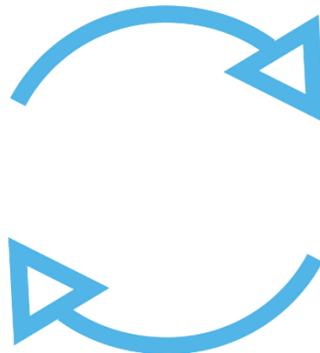
Where Does Administrative Data Come From?

Program start-up documents

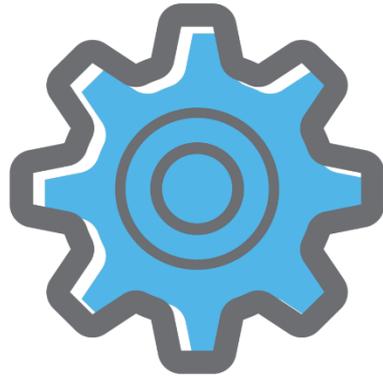**Grant applications**

Performance management metrics

Data collected for
- Case management
- Cost measurement
- Program auditing

Department of Labor (DOL) Enforcement Database provides public access to information collected about investigations conducted by the department's enforcement agencies, including the Wage and Hour Division (WHD) and the Occupational Safety and Health Administration (OSHA).

Department of Education, National Center for Education Statistics (NCES) provides public use access to a variety of programmatic data collections, such as the Integrated Postsecondary Education Data System.

What Kinds of Data Will You Release?

The kinds and formats of data you release will be guided by the reasons and the audiences for releasing the data:

Audience: General public or media?

- Consider basic program cost and performance data in aggregate form (e.g. tables of program outcomes by program sites)

## *Reason:* Support programmatic research?

- Consider detailed cost, performance, and outcomes information at a micro-level (e.g. a dataset of program site and outcome data by program participants)

Format: What type of files?

- Tabular data
- Micro-level data
- One-time data release
- Periodically updated data

First, Do No Harm to Mission

- **If an agency releases poor quality data, it may harm the agency's reputation and mission, rather than promote it.**
  - Attributes of poor quality data:
    - Contains errors and missing information
    - Lacks adequate documentation
    - Exposes the identity of respondents
    - Released in formats inaccessible to data users

- **Therefore, special preparation in releasing the data is necessary that goes beyond mere transparency.**
- **Agencies can avoid this risk by thinking carefully about a final question: *"How will you release the data?"***
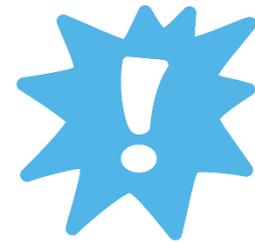
How Will You Release the Data?
Four Major Considerations
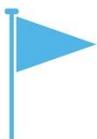
1. **Data quality**

   Only publicly release administrative data that has the smallest amount of missing information and the fewest errors possible

2. **Documentation**

   Provide enough information to allow data users to reproduce the analysis results and to accurately conduct further analysis.

3. **Disclosure risk and limitation**

   Work to identify and limit sources of disclosure risk in the data, especially identifiable information.

4. **Methods for accessing and using the data**

   Strive to release the data in formats and on platforms that reflect the technical capabilities of the intended audiences.

Listening to the Public

- **HHS Demand-Driven Open Data Project: http://ddod.healthdata.gov/wiki/Main_Page**
- **Demand-Driven Open Data (DDOD)**
  - DDOD is a framework of tools and methodologies to provide a systematic, ongoing and transparent mechanism for you to tell public data owners what's most valuable.
  - It guides Open Data initiatives by giving external users a way to ask for the data they need .
  - All work is entered, prioritized, implemented, and validated in the form of "use cases." This approach allows for all projects to have a known value even before work begins (Lean Startup approach to open data initiatives).

Can I Change the Question?

- **Asking about building demand for data may be the wrong question for Federal policy makers.**

- **A better question may be: "How can data owners make their data accessible without risking the Agency's reputation?"**

To continue exploring these questions,
read Summit's new white paper:

**"Are Your Administrative Data Ready for Public Use?"
By China Layne, Ph.D.**

Available at http://www.summitllc.us/APDU-2016

Team Contact Information

Anthony Curcio | Federal Credit Practice Lead

202-407-8303 | Anthony.Curcio@summitllc.us

Heather Brotsos | Manager, Applied Statistics

202.760.2457 | Heather.Brotsos@summitllc.us

China Layne, Ph.D. | Manager, Data Analytics & Research

202.407.8305 | China.Layne@summitllc.us