The Open Data Agenda Domain-Specific Discussions

As reported by: Theresa Pardo

Data Transparency Coalition 2013
September 10, 2013

The Task

- Recommendations to Congress and the Executive
- Three to five clear and measurable recommendations

Thanks to the Discussion Leaders!

- Daniel Schuman, CREW
- Kat Townsend, USAID
- Danny Goroff, Sloan foundation
- Haley Van Dyck, OMB
- Cinthia Schuman, Aspen Institute
- Tom Pollak, Urban Institute
- Jim Harper, Cato Institute
- Ed Walters, FastCase

Some comments on the whole

- New bodies
- New capabilities
- New assets and tools
  - laws; definitions; standards
- New requirements
- New funding
- Implementation recommendations

Group 1: Federal Spending

- Create a public advisory committee.
- Create a body to support cross-agency dialog of data processes and standards.
- Establish "terms of art".
- Require the use of consistent budget codes across government.
- Require the sharing of "actual" prices paid to vendors.

Group 2: Management and Performance

- **Management/Guidance:**
  - Add the need for "open standard" to any requirement of machine-readable format.
  - Who are the government POCs for the data, and what is their accountability? Right now the policy identifies a need for an email address of a data owner. Who would be the most useful POC for agency data?
  - OMB as maintainer of what a "program" means, and how this relates to PRA. E.G. "These are the programs we expect to see data out of for 53-300"
  - Interoperability of lexicon- what does a program entail for Treasury v. FCC v. DOS
  - What is the guidance for latency and timing of data releases?
  - Agencies need chain of business ownership that cannot be separated from IT ownership.

- **Reporting**
  - At the earliest stages of designing the implementation, ask who will use this information? Work with them at the onset to ensure the format and presentation is useful.
  - Outline requirements for a collection process and what terms mean. Train the implementers. Ensure everyone who touches the process understands the value, why the policy has been made, and how is it useful?
  - Need professional development/training of reporters, analysts, users, etc. to properly document and share data.

- **Engagement**:
  - Develop reward system for those that do use the data. Requires us to answer the question - How will usage be tracked?
  - Track questions made about the data once it has been made open. What are the cost implications?
  - Document the process of using and reporting data and what the burden is for the implementer. It may be far less than feared.

Group 3: Financial Regulation

1. Support the establishment and implementation of identifier systems that are public goods, including the global legal entity identifier system that the G20 is constructing. Open identifiers should track not just business card data, but also information about subsidiaries and hierarchical relationships, about contracts, etc. In particular, Congress should strike all federal requirements to use CUSIP, DUNS, or other proprietary systems. Instead, federal law should facilitate the use of free, open, and standardized systems instead. Other industries do not make you pay to use their parts numbers!

2. Expand XBRL tagging requirements to more SEC forms, including the Living Wills required by Dodd Frank.   Eventually extend more such reporting requirements to hedge funds, private equity, and the shadow banking system, too.

3. When opening data, establish public-interest advisory boards that can help set priorities, provide use cases, and devise data consumption tools that will seed the development of active user communities. Consider setting up an independent body modeled on the Open Data Institute in the UK.

4. Treat the reporting that the federal government requires regulated entities to do as open data, too. So instead of making a financial institution file similar information in different forms to different regulatory agencies, allow them to provide one set of open data and metadata that various regulators can download and use for their own purposes.

5. Strive for data quality by all means possible, including: pre-filing consistency checks, post-filing audits, easy mechanisms that allow users to identify and correct mistakes, greater speed in posting data, more metadata about provenance, simple ways of linking datasets to detect inconsistencies, etc.

Group 4: General Regulation

1. Make Federal Regulations Machine Readable by Default - Proposed Solution: Executive Order

2. All comments on federal regulations should either be submitted online OR made machine readable if submitted offline

   --Tied for third---

3. Develop Data Standards, Specifications, and ontologies for for federal regulations (for how to describe federal regulations) to enable
4. Publish the CFR in bulk, structured and machine readable format
5. Publish a comprehensive, structured Federal agency Org Chart

   ---Tied for fourth---

6. Tag each paragraph (or even sentence) with the Federal Register Notice that created it and with teh underlying statute that caused the Federal Regulation notice to be created.
7. Prohibit procurements and contracts that allow a vendor to privatize and resell govt data that ISNT open
8. Link/connect enforcement and reporting data to regulations (not just enable statutory provisions)
9. Common/standard language for health data sharing

Group 5: Tax

- Congress should require all nonprofits to file the Form 990 electronically and direct the IRS to release the data in an open format.

- Create a "safe harbor" for government data enabling it to release raw data without fear that it is not perfect.

- Promote uniform data standards across states and build form existing models, rather than starting form scratch.

Group 6: Legislation and the Code

- Open up committee data
  - Important work, lots happens here.
  - Membership, roles, calendar, hearing witnesses, testimony n digital form, meeting identifiers, synchronize between meeting records and videos.
- Open up Bill Data
  - A small but significant change would do wonders.
  - House/Sen calendars are good resources but still require scraping – down with scraping!
- Give us the Congressional Record in locator code!
- Create capability to actively track amendments to laws and to bills
  - This is hard!
  - Need unique IDs
  - Need both "cut and bite" version and red-line versions

Group 7: Judicial

1. Courts should not operate a walled garden for public domain judicial opinions. Congress should fund PACER out of appropriations and require PACER to be maintained as a single, open system, with the ability to download data for free and in bulk.

2. The Administrative Office of U.S. Courts should create and enforce standard data and metadata formats for judicial opinions within the federal court system.

3. The final, official versions of the law must be open. The Administrative Office of U.S. Courts should require, as a condition of private publishing contracts, that publishers provide back to the courts, free of charge, the official, paginated copies of judicial opinions in digital format. the AO should ensure that these opinions are included in the free and open PACER system recommended above.

4. The government should not create or enforce barriers to public domain law. Congress should establish that public law works of state governments, such as judicial opinions, state statutes, and regulations, are not protected by federal copyright law.

Some comments on the whole

- New bodies
- New capabilities
- New assets and tools
  - laws; definitions; standards
- New requirements
- New funding
- Implementation recommendations

The Open Data Agenda Domain-Specific Discussions

As reported by: Theresa Pardo

Data Transparency Coalition 2013
September 10, 2013