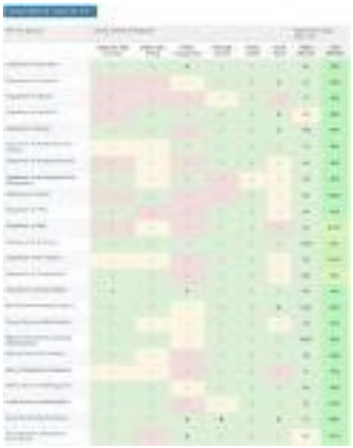
# Dashboarding Open Data Program Governance

Dennis D. McDonald, Ph.D.

# Dashboarding Open Data Program Governance

*May 20, 2015*

*By Dennis D. McDonald*[1]



An interesting aspect of NOAA's Big Data Project experiment is the "outsourcing" of selected NOAA data access and commercialization efforts to private sector cloud vendors including Amazon, IBM, Google, and Microsoft. Reliance on private sector partners for providing both public access and commercialization support raises the question about how NOAA -- or other Federal agencies using this approach -- will track program performance.

Some NOAA data sets that are potential interest to the public may already have specialized users both inside and outside the government. Such users (e.g., university based researchers) have already developed specialized access methods to obtain data feeds via custom connections with NOAA systems. Pushing such data out to "public" cloud resources could in theory make NOAA data resources even more available than such specialized access methods. How will NOAA know if this actually happens?

One possible reporting model is used by the Federal government's Project Open Data. This is governed by the Office of Management and Budget (OMB) and the Office of Science and Technology Policy (OSTP). This effort, assisted by GSA, provides both technical and policy guidance to ongoing efforts to "open up" government data across multiple agencies on a dataset by dataset basis.

One of the most visible components of this effort is the Project Open Data Dashboard. This is a top-level agency-by-agency report on progress being made in opening Federal agency data. Below is a snapshot of the current dashboard that is updated quarterly:

---

[1] *Copyright © 2015 by Dennis D. McDonald, Ph.D. Dennis is an independent management consultant based in Alexandria, Virginia. His experience includes consulting company ownership and management, database publishing and data transformation, managing the integration of large systems, open data, corporate technology strategy, social media adoption, statistical research, and IT cost analysis. Clients have included the U.S. Department of Veterans Affairs, the U.S. Environmental Protection Agency, the National Academy of Engineering, and the National Library of Medicine. He has worked as a project manager, analyst, and researcher in the U.S. and in Europe, Egypt, and China. His web site is located at* ***www.ddmcd.com*** *and his email address is* ***ddmcd@yahoo.com****. On Twitter he is* ***@ddmcd.***

Selected: Milestone 6 - February 28th 2015

| CFO Act Agencies | Leading Indicators Strategy | | | | | | Public Data Listing Metrics | |
|---|---|---|---|---|---|---|---|---|
| | Enterprise Data Inventory | Public Data Listing | Public Engagement | Privacy & Security | Human Capital | Use & Impact | Public Datasets | Valid Metadata |
| Department of Agriculture | ☑ | ☑ | ★ | ☑ | ☑ | ☑ | 505 | 100% |
| Department of Commerce | ○ | ○ | △ | ☑ | ☑ | ★ | 313 | 100% |
| Department of Defense | ○ | ○ | ○ | △ | ☑ | ○ | 372 | 100% |
| Department of Education | ○ | ☑ | ☑ | ☑ | ☑ | ★ | 275 | 100% |
| Department of Energy | ☑ | ☑ | ○ | ☑ | ☑ | ★ | 1026 | 100% |
| Department of Health and Human Services | △ | △ | ☑ | ☑ | ☑ | ☑ | 710 | 100% |
| Department of Homeland Security | ○ | △ | ○ | ☑ | ☑ | △ | 309 | 100% |
| Department of Housing and Urban Development | ☑ | △ | ☑ | ○ | △ | △ | 199 | 100% |
| Department of Justice | ○ | ☑ | ○ | ○ | ☑ | △ | 974 | 100%[1] |
| Department of Labor | ☑ | ○ | ○ | ☑ | ☑ | △ | 364 | 100% |
| Department of State | ○ | △ | ○ | ☑ | ☑ | ○ | 124 | 98.4%[1] |
| Department of the Interior | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | 79739 | 99% |
| Department of the Treasury | △ | △ | ○ | ☑ | ☑ | △ | 312 | 99.4% |
| Department of Transportation | ☑ | ☑ | ○ | ☑ | ☑ | △ | 3411 | 98% |
| Department of Veterans Affairs | ★ | ☑ | ★ | ☑ | ☑ | ☑ | 889 | 100% |
| Environmental Protection Agency | ☑ | ☑ | △ | ☑ | ☑ | ★ | 3674 | 100% |
| General Services Administration | ☑ | △ | △ | ☑ | ☑ | △ | 153 | 100% |
| National Aeronautics and Space Administration | ☑ | ☑ | △ | ☑ | ☑ | ☑ | 16629 | 100% |
| National Science Foundation | ☑ | △ | ○ | ☑ | ☑ | ☑ | 130 | 100% |
| Nuclear Regulatory Commission | △ | △ | ○ | ☑ | ☑ | ○ | 33 | 100% |
| Office of Personnel Management | ☑ | ☑ | △ | ☑ | ☑ | ☑ | 624 | 100% |
| Small Business Administration | ☑ | ☑ | ○ | △ | ☑ | ☑ | 36 | 100% |
| Social Security Administration | ☑ | ☑ | ★ | ★ | ☑ | ★ | 712 | 100% |
| U.S. Agency for International Development | ☑ | ○ | ★ | ☑ | ☑ | ☑ | 183 | 100% |

The rows in the table represent agencies submitting data to the dashboard. The right-hand column displays the number of data sets being reported on.

Each agency has a separate Dashboard page that summarizes data gathered via daily automated "crawls" of agency datasets. Crawls run every 24 hours with end-of-the quarter "snapshots" generated to reflect a past quarter's progress, as in the above sample. Agencies may also have their own more detailed webpages providing additional information on their open data program; for example, the USDA's own "digital strategy" efforts are reported here.

Dashboards like the sample displayed above combine manual, automated, quantitative, and qualitative data to measure progress. The color coding in the above example provides an indication of "leading indicators" showing if milestones have been reached that quarter, are in danger of being missed, or have actually been missed. While I'm personally not a great fan of such gross project measures, I am interested in the information provided by regular polling of agencies about their open date efforts:

- Are links working or broken? If they fail what error code is returned?
- Is the agency-reported inventory of data sets complete? For example, does a search for .CSV, .XML, or. JSON files show higher numbers than are being reported by the agency? If so, why?
- Are the means for the public to obtain contextual information about the data being provided?
- Are data owners and their contact information being provided?

Granted these types of dashboard figures don't give an indication of data set *usage* via the agency or via data.gov. They can provide insight into accessibility including efforts to standardize access, plus they provide useful practical information about data quality.

How might the above approach be relevant to something like the NOAA big data project or to other Federal agencies that are exploring private sector partnerships for improved data access?

One implication is that the data-issuing agency might want to negotiate agreements with private sector cloud vendors that incorporate some level of reporting and feedback to the source agency so that levels of data access can be tracked. Also, if different partner vendors make related data sets available, would it be possible to ensure that data and metadata standards are being implemented and maintained across data sets and by the participating partner vendors? If the same data set is being provided freely to the public and is also the basis for a commercial product based on further processing of the source data, would it useful for the source agency to know about such crossover?

One has to be careful about making such reporting requirements mandatory. Note the pushback that use of XBRL markup has received from some businesses involved in financial reporting.

A key feature of the Project Open Data effort being managed by OMB and OSTP is that so much of it is being conducted *in the open* using accessible resources such as shared documentation, a defined metadata schema, and use of GitHub for capturing comments and issues. Agencies that want to involve private sector vendors in their open date efforts should consider the use and management of such tools as a required part of program governance and oversight (as long as sufficient staff and resources are provided to manage such efforts, of course).

***Related reading:***

- *Breakthrough Financial Open Data Legislation To Be Introduced May 20*
- *The Knight Foundation's Civic Tech Report: "Open Government" Expenditures*
- *The Continuing Evolution of Data.gov*
- *Getting Real About "Open Data" Part II*
- *Observations and Questions about Open Data Program Governance*
- *Will NOAA's "Big Data Partnership" be a Model for Other Government Agencies?*
- *On Defining the "Maturity" of Open Data Programs*
- *Interim Report on the Generalizability of the NOAA Big Data Project's Management Model*
- *OMB Releases Federal Data Inventories – So What?*
- *Moving to the Cloud: Business as Usual or Opportunity for Change?*