Making Open Data Accessible Inside the Firewall

by
Dennis D. McDonald, Ph.D.[1]
November 26, 2013

Introduction

One of the most interesting sessions at last week's Nextgov Prime meeting in Washington DC was chaired by Saf Rabah, VP Products, Socrata. He was joined by Gray Brooks, Senior API Strategist, GSA, and Marina Martin, CTO, U.S. Department of Veterans Affairs. The subject of the discussion was the **provision of data access within the firewall**.

Discussions like this often focus on public access and "transparency." Here the panelists discussed what's involved in making data resources available within the organization.

Conceptually some of the discussion reminded me of the traditional justifications for "knowledge management" initiatives:

- Reduce duplication
- Minimize re-inventing the wheel
- Reduce costs
- Increase efficiency

Here are some of the takeaways I got from my note taking from the fairly freewheeling discussion.

1. Open Data is not always about transparency to citizens outside the government. It's also about making government more effective through improved access to useful data.
2. Creating an inventory and catalog of data is a good place to start. This forces a focus

---

not only on data descriptions but also on governance of the process by which decisions about data sharing -- with the public and with other agencies -- are to be made.

3. The cataloging stage is also good place in which to consider the type of metadata that needs to be generated to provide filtering and selection tools for users.
4. Agency size makes a big difference in the governance process. Larger agencies will will have more organizational "siloing" and potentially more organizational and political roadblocks to sharing.
5. Thought needs to be given to standards, tools, and the processes by which data are "cleaned up" and made ready for access and use.
6. We need to figure out what form governance and data standards should take. Should there be a "chief data officer"? Should there be a corporate "data architecture" to aid in planning? If so, what levels of authority and responsibility will be required?
7. You have to balance effective planning against the freedom to experiment and create. You don't want to create a bureaucratic process that stifles creativity.
8. Above all you need to focus on providing API's that give people the tools and data access they can then use to build their own applications based on real-world use cases that they define.
9. The "crawl/walk/run" analogy is a good one, keeping in mind that enabling self-service is a good goal. Starting out small to increase the likelihood of success is also a good idea especially when accompanied by linking up with a sponsor.
10. Whatever you do make sure the work is documented and sustainable so that people coming in from the outside -- e.g., from other departments or agencies -- know what they can expect to get.
11. People needing access to Federal data, particularly the states and local governments involved in a program, appreciate a single data source.
12. Keep in mind that data sources that target external users (e.g., Data.gov) are also heavily used by internal people.

Discussion

As noted above this Nextgov Prime session was one of the better discussions of this topic I've heard. The focus on data access "within the firewall" was much appreciated. In my own white paper [A Framework for Transparency Program Planning and Assessment](#) I was adamant about the need to focus on the needs of the agency that generates the data not just on the needs of external users and hoped-for innovative and commercial uses; these were my recommendations and I think they align well with what the three Nextgov Prime speakers discussed:

1. Don't spend money on converting or standardizing data that are old or out of date.
2. Start small so that the initial "humps" of data conversion costs and having to operate dual or overlapping systems won't sink the program.
3. Align transparency activities with the programs responsible for supplying and using the data.

4. Make sure that leadership, operations, and budget are adequate and reliable.
5. Target critical use cases that generate real benefits to users.
6. Avoid an "if we create it they will come" strategy.
7. Make sure all stakeholders are on the same page — agency staff, procurement staff, legislators, non-governmental participants, and most importantly, users.
8. Focus program metrics on outcomes, not just internal transactions, costs, and increased program efficiency.
9. Target mobile devices first for exposing data to different constituencies through creative methods of data visualization and manipulation.
10. Maximize reliance on open source, non-proprietary, and off the shelf tools and techniques.

The importance of governance

Of all the points made by Rabah, Brooks, and Martin, the most important were those related to **governance**. Too much and you stifle innovation and creativity. Too little and you lose the advantages of scale and standardization.

A governance process needs "teeth" to provide meaningful leadership and accountability. It also requires resources and authority to manage all the different activities associated with what I'll the "data access lifecycle," starting with a definition and understanding of the data access program's goals and objectives, all the way through development, implementation, and operations.

Failure to provide necessary governance -- and leadership -- is one of the reasons it's taking so long to provide standardized access to Federal financial information, as discussed in [A Project Manager's Perspective on the GAO's Federal Data Transparency Report.](#) Oversight and coordination are no substitute for management, especially when we are dealing with the almost bewildering array of data files of all types that can be revealed through a serious data inventory.

Balance between control and autonomy will be critical and will vary by program size, complexity, structure, and maturity. Also critical will be how external and internal users of related data resources are supported when important distinctions related to privacy and security need to be maintained. This will become increasingly difficult as personal, mobile, and cloud-based technologies continue their adoption as authorized work tools.

Related reading:

- *[A Framework for Transparency Program Planning and Assessment](#)*
- *[Questions About Google and Local Government Data](#)*
- *[The Changing Culture of Big Data Management](#)*
- *[AGA's Citizen Centric Reporting and Government Transparency](#)*

- *How To Make Datathon Efforts Sustainable*
- *Planning for Big Data: Lessons Learned from Large Energy Utility Projects*
- *When Federal Government Project Management Collides with the "New Normal"*
- *A Project Manager's Perspective on the GAO's Federal Data Transparency Report*
- *Understanding How Open Data Reaches the Public*
- *The State of Government Data Transparency, 2013*