



# **Observations and Questions about Open Data Program Governance**

Dennis D. McDonald, Ph.D.

---

# Observations and Questions about Open Data Program Governance

February 27, 2015

By Dennis D. McDonald<sup>1</sup>

## Introduction

My interest in open data is influenced by involvement with many types of data access projects ranging from market research surveys, illustrated parts catalogs, and cancer research publishing to customer relationship management.

After this week's call about working with data owners<sup>2</sup> I put together a few thoughts to share with the group. Here are some observations along with a list of questions that I think deserve further discussion.

Topics include:

1. The importance of stakeholder involvement
2. The impact of resource constraints
3. Controlling business process change
4. Importance of the data management lifecycle
5. The importance of use cases
6. The role of subject matter expertise

## 1. The importance of stakeholder involvement

From a project management perspective everyone will agree with the central importance of stakeholder involvement. The trick is, in open data programs, who are the stakeholders? Once you've identified them, how do you engage with them?

---

<sup>1</sup> Copyright © 2015 by Dennis D. McDonald, Ph.D. Dennis is a 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, 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 throughout the U.S. and in Europe, Egypt, and China. His web site is located at [www.ddmcd.com](http://www.ddmcd.com) and his email address is [ddmcd@yahoo.com](mailto:ddmcd@yahoo.com). On Twitter he is [@ddmcd](https://twitter.com/ddmcd)

<sup>2</sup> This article is based on the author's participation in an online discussion among local, state, and U.S. national health data publishers on February 24 sponsored by Socrata Inc. The views expressed here are those of the author and are not intended to represent the official views of Socrata or of any of the other participants.



One thing that sets open data projects apart is that internal stakeholders may be spread throughout the organization. External stakeholders such as end-users and intermediaries may be difficult to reach. Identifying and engaging with stakeholders will take time, especially when they control the data and systems that generate open data assets.

Key questions:

- a. How do we identify stakeholders?
- b. How do we engage with them even if they are dispersed throughout the organization?
- c. What roles do we assign them in the overall management of our open data program?

## **2. The impact of resource constraints**

Not everyone has the luxury of having a fully funded and independently managed open data program. Many require involvement by staff on top of everything else they're doing. This reality emphasizes the "grassroots" nature of many open data project which, on the upside, can tap into energy and involvement of committed staff.

The downside is that, when a complex project has to be helmed by part-time staff, schedules, quality, and cost control can suffer.

Key questions:

- a. What kind of a management approach is appropriate to controlling a project employing "bootlegged" staff versus one employing full-time or dedicated staff?
- b. How do you keep track of costs, hours, and other resources?
- c. When is it appropriate to involve volunteers or civic activists in key project roles? How do you manage them?
- d. When is it not appropriate to involve volunteers or civic activists?

## **3. Controlling business process changes**

Inevitably the creation and management of an open data program will require someone to change how they do things. This can be a source of resistance. Motivating people to change how they work becomes necessary but may be a challenge in situations where top-down leadership is lacking or when potential open data assets or silos are independently managed. Also, saying that "business processes and the data and systems that support them will have to change" presumes that an understanding of "current state" and "future state" exists.

Key questions:

- a. How can we "sell" business process owners on changing how they operate in order to participate in the open data program?
- b. When is it appropriate -- and possible -- to minimize changes on the part of data asset owners?
- c. How can we help participants to manage how they change their business processes?

## **4. Importance of the data management lifecycle**

Each data element and its associated metadata we're thinking of making public needs to be managed from the time it is created through the time it's retired. That's what is meant by "data management lifecycle."

Ideally those closest to the systems where data and metadata are generated and controlled will be better positioned than others to understand the technical implications of making these data public. At the same time, those technically qualified to understand technical data management may not be the most qualified to understand the needs of all users and the uses they intend to make of open data.

Key questions:

- a. What are the critical first steps in implementation of a data management lifecycle approach to supporting an open data program?
- b. How do you manage the relationship between a data management lifecycle process and the programs that are currently producing and using the data that will be made "open"?

## **5. The importance of use cases**

Understanding user requirements has always been an important component in system development. That's one of the reasons people do market research and user requirements research at the start of a development effort.

The first generation open data programs focused on making data files available via a catalog type interface. This is evolving as people move beyond a "transparency" focus to ensuring that open data programs also deliver useful data in a usable fashion so that positive impacts can be generated from consumption of the open data.

Focusing on technology and the size and number of data files made available via a portal need not distract from a focus on users and data usage. One can start by creating a well-documented set of "open data use cases" and then work back to understand what types of process, system, and data changes will be needed to support those use cases.

Key questions:

- a. What, if any, barriers exist to adopting a use case approach to documenting what an open data program should be accomplishing?
- b. Use cases can focus on "what is needed" not just on "what's available." Is there a downside to concentrating on a few high impact use cases instead of less complex use cases that address a high proportion of the same data that are potentially available? (For example, deliver less data but of higher value?)

## **6. The role of subject matter expertise**

One of the first questions to address in the design of an open data program concerns the type of data the program is going to deliver and to whom.

There are many ways to classify data: topic of program area, function, physical format, animal versus human, numeric versus graphic, etc. Programs that focus on delivering health services, for example, can generate data describing service recipients, nature of service delivered, diagnosis, cost, treatment details, geographic location, institutional affiliation, etc.

It's often necessary to involve subject matter experts and open data program design and operation since they're professionally qualified to understand the unique nature of the systems and processes involved as well as the language and metadata associated with the systems and processes. At the same time, there may also exist areas of the open data program where subject matter expertise is less important than an understanding of more generic management and technology concerns.

Key questions:

- a. In what areas of open data programs is subject matter expertise critical?
- b. In what areas is it less important?
- c. How do you make the distinction?

### ***Related reading:***

- [\*CityGram NYC: A Model for Open Data Efforts\*](#)
- [\*Compendium: My Guest Posts for the BaleFire Global Open Data Blog\*](#)
- [\*Don't Let Tools Drive Enterprise Data Strategy\*](#)
- [\*Getting Real About "Open Data"\*](#)
- [\*Getting Real About "Open Data" Part II\*](#)
- [\*The Importance of Audience Research to Open Data Program Success\*](#)
- [\*Management Needs Data Literacy To Run Open Data Programs\*](#)
- [\*OMB Releases Federal Data Inventories – So What?\*](#)
- [\*Open Data and Performance Measurement: Two Sides of the Same Coin\*](#)
- [\*Possible Pitfalls in Comparing Performance Measures Across Programs\*](#)