

# The Changing Culture of Big Data Management

by  
Dennis D. McDonald, Ph.D.<sup>1</sup>  
October 29, 2013

## Introduction

Mike Barlow's ebook [The Culture of Big Data](#) is a short and sweet introduction to the importance of approaching "big data" projects differently from the more highly structured data intensive systems and projects we may already be familiar with. Not only is the technology different, Barlow says, but the required skills, business processes, and governance are also different.



Kathy O'Neill says in her ebook [On Being a Data Skeptic](#) that the technology associated with big data can seem like mumbo-jumbo to business folks and that management and analysts need to do a better job of intelligently collaborating.

## Success elements

Barlow emphasizes the importance of starting with the problem not with the technology. Implementing an effective "big data" strategy has more to do with solving real problems, Barlow says, in ways that contribute value through increased profit, revenue, or efficiency. Barlow bases his book on a series of interviews with corporate "big data" practitioners. All seem to emphasize the importance to success of strong executive support and close attention to solving important enterprise problems. There is nothing "pie-in-the-sky" about these success elements. You could probably say the same about just about any new technological innovation and how it becomes useful in the real world.

## Focus on data

One of the things that separates "big data" initiatives from more traditional technology-based

---

<sup>1</sup> Dennis D. McDonald, Ph.D. is an independent consultant based in the Washington DC area. He has worked throughout the U.S. and in Europe, Egypt, and China. In addition to consulting company ownership and management his experience includes database publishing and data transformation, integration of large systems, corporate technology strategy, social media adoption, statistical research, and IT cost analysis. 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).

initiatives is the focus on data. This seems obvious but has important implications for managing initiatives. When you think about how we use computer hardware and software you think about how data are collected, organized, managed, and used. Data in many different forms are organized and manipulated in ways that support knowledge transfer and decision-making. When we talk about “big data” we’re talking about manipulating and managing data in volumes and variety that are potentially much greater and different than what we’ve been accustomed to dealing with in the past.

### **Focus on solutions**

Even if we address the challenges of increased data variety and volume, we still have to address whether or not we know how to take advantage of the patterns and inferences we derive from our analysis and modeling activities. Barlow suggests, reasonably enough, that it’s smart to focus on solving practical problems such as increasing revenue, increasing efficiency, and reducing costs.

It’s hard to argue with that. In most business situations prudence usually dictates focusing on real immediate problems and “pain points” to demonstrate value to management.

### **What about innovation?**

Balance is needed. We don’t want to be so “tactical” that we discourage potentially useful innovations from emerging.

This is one of the areas where I think Barlow falls short. On the one hand, analyzing “big data” can enable us to detect patterns and relationships that we might not otherwise see compared with our current market research or planning efforts. On the other hand, even if we detect new patterns, will we be in a position to take advantage of our new insights?

Hitching our corporate star too narrowly to solving currently perceived problems has the potential, if we’re not careful, of limiting the value of what we might derive from our big data initiatives. After all, too narrow a devotion to what the customer wants can, in the short-term, blind us to what the customer really needs.

### **Being a “data skeptic”**

O’Neill’s approach is to point out the possible pitfalls of relying too unquestionably on data analysis and modeling efforts that are not carefully thought out or whose limitations are not explained in understandable terms.

As an analyst she understands that “the devil is in the details.” She points out issues with measurement and causality that can be overlooked by those who focus more on the end results of analysis and modeling than on its internal plumbing. Having been in the technology

business for several decades now even I can smell the hype associated with selling new tools. Being able to manage and take advantage of them requires management effort starting with a few key points outline below.

### ***1. A basic understanding is needed***

Hiring a crew of “data scientists” to manage and model data does not absolve you of responsibility for understanding the basics of what it is they do. If you really want to manage something you need some level of understanding of what it is you’re managing.

O’Neill suggests that one important place to start is to appreciate what can actually be measured and what can only be approximated. Having designed many data collection processes I can personally vouch for this. Are you measuring actual behavior or activities where data and their meaning are unambiguous, are you basing your data on personal reporting that requires potentially faulty memory, or are you using secondary or tertiary proxies for what you are trying to predict?

You don’t have to be a data scientist to understand the significance of these questions and you should be able to get straight answers about the data from your data people.

### ***2. Common sense is required***

Don’t let data models preclude common sense. If the numbers that pop out of the analysis are too good, be skeptical. And if the numbers are telling you something you don’t really want to know, don’t ignore them. Find out why they’re bad and what that really means to your business — and tell management about it.

### ***3. Complexity is real***

Don’t assume that the results of a complex analysis can be completely described in a single PowerPoint slide and a few bullet points. As O’Neill suggests, bring your analysts into your strategy meetings. Not only will this help them understand the business implications of what you are planning but they will also be able to shed light on the reality — or unreality — of your plans and assumptions.

### ***4. What do you mean, “all”?***

It may be true that “big data” approaches can provide significantly more data for analytics and modeling than traditional sampling or surveyed based approaches. In those approaches inferences about a particular population were based on small numbers of observations carefully selected for “representativeness” and predictability of results. With today’s big data approaches it is attractive to think that we have “all the data” to work with and that using modern data architectures and analysis tools we can base our analyses on the total population.

Still, we have to be careful about that word “all.” Yes, we may have sales data for “all” our customers and an associated set of customer characteristics for them. But what if our real questions require us to know whether or not our customers are actively representative of the general population? Are we using as part of our analysis general population data on, say, homeownership or debt as a proxy for the homeownership or debt characteristics of our own customers? We need to be careful about what that word “all” means and how it relates to our underlying analysis assumptions.

## ***5. Data are neither good nor bad***

O’Neill makes this point explicitly: data are neither good nor bad. It’s how you use them that makes them seem one way or the other. Perfectly clean, sound, accurate, up to date, and all encompassing data can be put to nefarious uses.

Does it work the other way, i.e., can good people make good use of poor data, where poor is defined as inaccurate, incomplete or nonexistent?

I would argue that the answer this question is “yes” but only if the people, processes, and tools and how they interact are open and transparent with each other about the limitations of the sources. Analysts and business people have to be open and honest about how such data are used and how “flaws” might impact error rates, risks, or uncertainty. Creative collaboration really helps when dealing with the realities of real world data.

## **Conclusions**

In some ways managing “big data” tools and processes is no different than figuring out how to manage any other type of technological innovation. The technology is introduced, experts emerge and help control and shape evolving practical applications, and management eventually figures out what is worth keeping and what can be discarded. Barlow and O’Neil, by focusing on people and processes, help move us along along this still evolving path.

Yet, big data does offer challenges that many analysts and managers are going to have difficulty reconciling. Analysts really do need to understand more about business and business strategy than might have been the case in the more compartmentalized past. At the same time, managers who don’t understand and appreciate how data analysts work and how trends, modeling, and error are handled will be at a disadvantage. The two groups need to work together to make “big data” work.

## ***Related reading:***

- [\*Planning for Big Data: Lessons Learned from Large Energy Utility Projects\*](#)
- [\*A Project Manager’s Perspective on the GAO’s Federal Data Transparency Report\*](#)

- [\*The State of Government Data Transparency, 2013\*](#)
- [\*More Data, Please - and Don't Hold the Pickle\*](#)
- [\*Learning from the World Bank's "Big Data" Exploration Weekend\*](#)
- [\*Recouping "Big Data" Investment in One Year Mandates Serious Project Management\*](#)