

**The DATA Act:
Good Use of Scarce Government Resources?**

*Closing Keynote: DATA Act Summit
Data Coalition, Washington DC
Washington Hilton
May 26, 2016
Shelley H. Metzenbaum*

The Opportunity

Data. Analytics. Big Data. Data-driven.

More and more people are talking about data and analytics these days. Even Congress joined in when it adopted the cleverly named DATA Act.

There is good reason for so much attention to data.

- Data and data analytics can *reveal problems* and *opportunities for improvement*.
- They can inform *priority-setting* and *program design*.
- They provide a *quick read whether or not government actions are having the anticipated effect*.
- Good data sets, plus data from controlled trials, *help determine if and how much government actions explain changes in performance trends*.
- They can also suggest *ways to reduce costs without compromising results*.

Data is *only* valuable, however, when wisely collected, analyzed, shared, and used to find ways to improve on multiple dimensions.

The Challenge

Today, I want to talk about data and the DATA Act.

Specifically, I want to encourage everyone in this room to do whatever you can to make sure the reporting and related requirements of the DATA Act realize more value than they cost and more value than alternate uses of the same resources. Today, I want to urge each and every one of you to start to ask, answer, and act on the following question:

What needs to happen to make sure the work being done to comply with the DATA Act adds significant value to *government's ability to deliver more mission for the money in fair, understandable ways that also increase trust in government?*

Isn't that the question we should be asking before, during and after implementing this and any law?

Value of Data and Analytics

I am a big fan of data and a big fan of analytics. In its many forms, both are critical to *more effective, efficient, fair, understandable, and trusted government*. I am not, however, a big fan of generating data without a clear purpose.

Wise handling of data can and has changed the world. Consider, for example, the high value of the data federal, state, and local governments collect, share, and use about harmful incidents such

as mortality, morbidity, traffic fatalities, airplane accidents, chemical spills, and fires. The data sets that have been assembled to *prevent* and *reduce* the *consequences of these incidents* reflect careful and continuing thought about the *kinds of questions these data can answer* and the *dimensions of collected data* needed to generate those answers. These data are useful not just for a central office, but to the front-line and across the delivery chain. Traffic fatality data, for example, are useful not only to federal and state policy and budget decision-makers, but also help my local public works department decide which intersections are the most dangerous and should be considered as priorities for redesign. They also help local government learn from the experience of other communities that have faced similar situations.

Let me explain what I mean by the dimensions of the data.

- Many incident reporting systems time-stamp their data, noting *when an event occurred* – the time of day, day of week, and week of year – and *how long the event lasted*. This information enables a picture of the *historic frequency, duration, and timing* of events, and aids the *discovery of the causal factors government can try to influence to prevent and reduce the consequences of harmful incidents*.

In addition, studying time-linked attributes can *trigger new ideas for more effective government action*. That is what one office of the U.S. Coast Guard discovered a number of years ago when it realized that most oil spills took place in the wee hours of the morning while all inspections occurred during working hours. Shifting inspection times to the wee hours resulted in fewer oil spills.

- Capturing *other characteristics of harmful incidents* in databases, such as their *location, operator and equipment characteristics, and other causal factors* enables *design of more precise, tailored, and, consequently, effective and cost-effective* interventions. How many of you have heard of the [Haddon Matrix](#)? If you have not and your organization wants to prevent bad things from happening and keep costs as low as possible when they do, take a look. It may help you think about how you want to count and characterize the incidents you are trying to reduce.
- Including information about the *cost of an incident* enables an agency to consider whether *lower-frequency, high-consequence events* warrant more attention than *more numerous, less consequential events*.

Of course, well-designed data *collection, analysis, and dissemination* are needed not just for harmful incidents, but also for beneficial events and conditions. Government collects reams of data about all sorts of events and conditions. Census, economic, labor, weather, and human genome data come to mind. When careful thought has gone into not just the *generation and collection of the data*, but also their *analyses and dissemination*, these data sets reveal opportunities, support a thriving market economy, inform personal choices, and encourage discovery, among other uses. The intense demand for data and reports from the Census, Bureau of Labor Statistics, and Bureau of Economic Analysis suggests the value of well-designed data management.

Other federal agencies work hard to manage their data in ways that make the information more useful to target audiences who can play a constructive role bringing about needed change. One of my favorite examples is the campaign material the National Highway Traffic Safety Administration (NHTSA) developed to increase the use of safety belts in autos and other vehicles to reduce traffic fatalities and their consequences. After analyzing traffic fatality data and seeing reduced fatalities in states using similar practices, NHTSA created the “Click It or Ticket” campaign to encourage adoption of those practices and developed and distributed campaign materials for targeted users, such as school principals and local police, to adopt and, if needed, adapt.

Unfortunately, attention to the users and uses of data and analyses is too often more the exception than the rule. Many federal [grant programs](#), for example, collect reams of data from grantees but fail to return that data to data suppliers and others with value added through analysis. Too few undertake user-centered design when they set up their reporting systems. [One grant program](#), for example, requires grantees to report on 1400 different performance standards. Unfortunately, no one currently analyzes the submitted information, nor returns analyses to grantees in a format that enables them work smarter. Fortunately, this federal agency is beginning to work with its grantees to fix this problem.

“Moore’s law,” the name given to Intel co-founder Gordon Moore’s forecast that computational processing power would double every 2 years – a phenomenon that has happened for the last 50 years – means we live in a very different world today than we did 10, 20, 30, and 40 years ago. We use data and statistical analysis in our lives far more than ever before. Sports teams, for example, use sophisticated Moneyball methods to identify better players and teams of players. Google and Mapquest analyze and return data that help us get to our destination. Even complex statistical concepts are working their way into our vocabulary and understanding. Today, for example, many more people understand what a false positive is than did 20 years ago, thank, I suspect, to home pregnancy tests.

These developments make it possible for us to *act*, and *interact*, in wholly new ways. They open up unprecedented opportunities to deliver more effective, efficient, fair, understandable, and trusted government.

In truth, though, smarter collection, analysis, and sharing of data don’t depend on Moore’s law. Way back in 1854, John Snow mapped houses in London with cholera deaths and saw they clustered around a single water pump. Even back then, without any computers, he was able to *collect, analyze, display, and share* data in a way that enabled people to see the action needed to slow the spread of cholera, in this case removing the handle of the pump dispensing contaminated water, thereby compelling people to gather water from other, non-contaminated pumps.

In short, effective data management is not just about technology. It is about a will, an attitude, skills, and actions that encourage data use and communication. Are data collected primarily to satisfy reporting requirements and vaguely defined notions of accountability? Or, are data systems designed to support *discovery, priority-setting, problem-solving, and continual improvement along multiple dimensions*? Isn’t it time to focus on designing and delivering data and analyses so they support continual learning and on-the-ground improvement, and where relevant, inform individual choices, instead of placing primary emphasis on compliance with legislated deadlines? Isn’t it time for oversight efforts to direct more of their attention to the value proposition?

What is the Expected Benefit of the DATA Act: Value, Uses, and Users?

You may be wondering why I am talking about the value, uses, and users of all these different kinds of government-collected data. You may be thinking. “Wait a minute, you are talking about the power of mission-linked data sets, not spending data. Spending data are different.”

But are they? The private sector has certainly figured out how to analyze our spending data to get us to buy more, although I doubt that is what we want the federal government to do.

The question is: how do we want federal spending data to be used? What is the purpose, the hoped-for benefit, of collecting the data required by the DATA Act? Who are likely users, what

are their likely uses, and how can those uses ultimately lead to better, more trusted government than would have occurred without the DATA Act?

When Hudson Hollister first invited me to address this Summit, I asked if he really wanted me to talk about the DATA Act because I could not figure out what it was trying to accomplish. Hudson assured me, with a chuckle, that this would not be a problem since Ralph Nader filled the same time slot on last year's program!

So, here I am, with lots of questions about the value you want the DATA Act to achieve, but not many answers. I am here with lots of fears about the value of the DATA Act, but at the same time, a hope and optimism, especially after what I have seen today, that by articulating those fears, you will work to make sure my fears are not realized. And I am here with some suggestions about implementing and conducting oversight of the DATA Act, with the hope you will indulge me if some of those ideas don't make sense and let all of us know what does.

Let me start with questions about the *value* of the DATA Act.

The DATA Act identifies as one of its purposes that it should:

- enable taxpayers and policy makers to *track* Federal spending more effectively

I have seen a lot of tracking that yields little useful information so would like to understand, more explicitly, what additional information you hope the DATA Act tracking will yield. I understand it can help government detect and reduce fraud, but what else do you want to learn and how will that information enable government or others to make better decisions and take smarter actions? If you want to meet the purpose of tracking federal spending more effectively, it is essential to be clear what you hope to accomplish with that tracking.

The second purpose of the DATA Act is the desire for:

- *consistent, reliable, and searchable* Government-wide *spending data*

It is hard to argue against more consistent, reliable, and searchable data, but I can already search government-wide spending on USASPENDING.gov. It would therefore help me, and I suspect others, to understand more specifically the consistency, reliability, and searchability problems supporters of the DATA Act want to see fixed.

Beyond that, establishing baseline indicators for key attributes of interest would help gauge the impact of the DATA Act, going forward. After what I learned earlier today about the responsibilities given in the law to Inspectors General (IG) to do statistical sampling to assess *completeness, timeliness, quality, and accuracy* of the data submitted under the Act by each federal agency, let me urge timely implementation of this provision as soon as possible but definitely *before* May 2017 when agencies are expected to start reporting under the new DATA Act requirements. This would afford a viable mechanism for this sort of assessment.

I love the third and fourth purposes of the DATA Act:

- *simplify reporting* for entities receiving Federal funds and improve the quality of data submitted to USASPENDING.gov.

In truth, before this event, I did not understand how the DATA Act would fix this. I am beginning to appreciate, though, that implementation of the DATA Act may:

- *link management and financial data*, making it easier to see spending status of appropriated funds; and

- *eliminate duplicative transactions reporting*, speeding access to and improving the quality of transaction information.

In short, I have long understood the “what” of the DATA Act, as I know others have, but until today, I could not understand the “why.”

Let me suggest that advocates, Congress, GAO, and the Executive Branch take time to answer and better articulate these “why” questions and lay out the theory of change and the evidence behind it in a more disciplined way. Further, let me suggest being more clear about the anticipated value, the likely users, and the possible uses.

My Fears

If that clarity of objectives and means to measure progress on them does not happen, I fear that GAO, Congress, and Inspectors General will simply monitor the status of the law’s implementation, but not its impact. Moreover, without a clear line of sight to the objectives, I fear that, in their oversight, these oversight bodies will not afford Treasury, OMB, and federal agencies the time to *test*, *assess*, and *adjust* different approaches to implementing the law to find the best way to accomplish its underlying, albeit not fully articulated, intent.

Let me list my specific fears with hope that spotlighting them may keep them from becoming reality.

1. **Lots of work with little payoff.** How will the DATA Act improve fraud detection beyond what USA Spending and individual programs, such as Medicaid, can already do? Is the marginal value of the DATA Act really worth its marginal cost? If not, how can adjustments be made? I must admit I have less worry about that now than when I was preparing my remarks after I learned, at today’s event, about possible benefits of the Act. In addition to the two mentioned above, these seem to include:
 - Creation of a *unique entity identifier*, if it helps other federal programs see non-compliance across programs or make progress creating unique identifiers allowing that;
 - *Easier grant filing* using pre-populated fields; and
 - *Lessons in agile systems development for other complex, government-wide data systems.*

Still, it would be good to articulate, track, and report progress on the full set of objectives so others can better understand what the DATA Act is trying to accomplish and, if needed, adjustments can be made to the objectives or strategies being used to make progress on them.
2. **Sucking the oxygen out of the room.** A lot of very smart, capable people around government are spending a whole lot of time these days implementing the DATA Act rather than applying their analytic, computational, and contract management prowess to other equally and arguably more serious problems, such as cyber-security, as well as to growth opportunities, such as finding mission-focused performance and productivity improvements, including helping grantees learn from and apply the lessons of their own and others’ experience. These kinds of skilled resources are scarce across government, and Congress needs to pay more attention not just to the budget cost of each new piece of legislation, but also to the opportunity cost.
3. **Hyper-emphasis on improper payments, error rates, and failures.** Reducing improper payments is important, but not at the expense of performance and productivity improvements.

Yet, over and over again, I have seen improper payment targets and improper payment penalties overwhelm agency and grantee attention to making progress on a program's purpose. I fear this system will exacerbate that imbalance, especially after hearing one of the IGs indicate, as a parenthetical aside, his discomfort with anything but a zero error rate. I especially fear this imbalance because I see so many IGs, auditors, and accountants in this room, but have yet to run into a program manager. Keep in mind that there are a number of circumstances where higher error rates may not be a huge problem. In setting up a new data system to measure mail delivery times, for example, former Postmaster General Marvin Runyon recalls his staff warning him that there could be a forty percent error rate. He said, "That's fine. It will at least show me where there is a really bad problem, and we can go to work on that."¹ Data used to trigger focused follow-up questions can tolerate a higher error rate than data used for enforcement actions or as the basis for incentive payments.

4. **Limited tolerance for iterative learning.** In GAO's latest review, it expresses concerns that revisions in technical schema increase the likelihood that agency implementation will be late. I have a different concern: that the press to meet deadlines will lead to the adoption of clunky, possibly wrong-headed system design choices. Speed is important, but, in this case, value is more important. I was delighted to see that Treasury and OMB, working with the 18F folks, are embracing user-centered, agile development. I hope, going forward, GAO and Congress conduct their oversight in a way that encourages aggressive, rigorous, *agile testing, assessing, adoption, and adaptation* to advance the intent of the law rather than focusing on simple compliance with the letter and deadlines of the law.
5. **Overly rigid definitions in some areas can encourage rigid practices.** In one of its oversight reports, GAO expressed concerns about definition variability over time, such as for program activity. If we want agencies to achieve better results on multiple dimensions, we want them to be able and eager to adjust their activities over time, dropping those that don't work and testing and adopting new activities to find those that work better. This is likely to result in a change in program activities, which should not be considered a problem.
6. **Congress will cull for problems to attack.** Let's be honest. Members of Congress get more media attention when they criticize than when they praise a federal agency, and media attention helps them get re-elected. Yet, if Congress focuses primarily on the problems and not the progress, it will frustrate and discourage the people doing the work. Is constructive Congressional oversight possible or just a pipe dream? I hope so, but what will it take us to get there?

My Hopes: Ideas, Questions, and Suggestions

So, I have shared my fears. What are my hopes and suggestions?

I will restate it. My hope is that, after the law is implemented, government is more *effective, efficient, fair, understandable, and trusted*.

To get there, let me offer a suggestion on the DATA Act.

Benchmark. I know the Treasury team has done this, but if you have not, I urge you to look at other government spending sites to learn from their experience. Maryland and Ohio both have spending sites, for example.

- List what you learn from each site and what you like and list what you did not.
- Figure out what problems the site might address or identify.

- Try to figure who is using the data or who might use it, for what purpose, whether they are in fact doing so, and changes in the world arising because of the site.
- After looking at a few sites, try to list the objectives you hope the DATA Act will advance. Do this whether you are Executive Branch, an advocate, GAO, Congressional staff, or even the media. The list of objectives is likely to evolve over time, which is fine. But, please, start to articulate not just the “what,” but also the “why,” of the DATA Act and how it is expected to improve USA Spending.

I looked at the Maryland and Ohio sites and found myself thinking about a few possible functions, other than fraud detection.

- **Vendors – who are they and how much government spending do they receive.** Certainly, the list of the largest vendors is interesting. Frankly, I found it especially interesting in Ohio, where one of the largest vendors is the Shelly Company, an asphalt company, but that is only because my name is Shelley. Still, competitors, investors, and even the vendors themselves, let alone those worried about fraud, might like to know how much money they have and are likely to get from the government.
- **Where dollars are spent.** The location of dollars outlaid might also be interesting from an economic development perspective, although I don’t think the DATA Act will be granular enough to provide that value despite some people’s expectations that it will. Before I joined the Obama Administration, I was at UMass Boston, where we expected to play a role in “Recovery Act” implementation, helping small local governments manage the funds they received. If (or when) that eventually happened, post DATA Act spending might show the dollars going to Boston, when, in fact, they were being spent in the smaller communities outside Boston. Tracking dollars to where they are actually spent would require tracking not just to the grantee and sub-grantee level but also to their grantees, contractors, and sub-contractors. This kind of tracking could prove very costly, and certainly burdensome, for grantees. Congress, GAO, and others need to be realistic in their expectations about the information the DATA Act will yield.
- **When dollars are spent.** Arguably, the speed and timing of spending might help fund receivers improve the *predictability* of revenue and Treasury improve its assessment of when it needs to borrow. Is that a need? Who could use this sort of timing data and will analysis of data collected under the DATA Act be able to produce it, especially without appropriation dates in the system?

I am hopeful that there is gold in ‘dem ‘der DATA Act hills. Data is powerful when useful and routinely *used* to make better decision that *improve government’s impact, efficiency, fairness, understandability, and trust*. If it is not, it is wasteful!

I urge all of you to, please, take the time not just to comply with this law, but to figure out the value proposition and how best to realize it.

Let me also offer a plea to advocates and Congressional staff in the room. Before proposing and adopting a new law, figure out first the specific problems in need of fixing or opportunities being missed. Be clear about the theory of change and evidence supporting the theory. If there are legal barriers or if a law can help make needed changes stick, move a law to fix the problem. If not, though, figure out instead how to conduct constructive oversight that will motivate the changes needed.

ⁱ Memorandum re “GET RESULTS THROUGH PERFORMANCE MANAGEMENT” to Government Executives from Executive Session on Public Sector Performance Management, Harvard University, Kennedy School of Government, 2001, available at <https://www.innovations.harvard.edu/sites/default/files/7101.pdf>