journal

OF GOVERNMENT FINANCIAL MANAGEMENT®

FALL 2017 VOL. 66, NO. 3

Protecting Governments' Data

AGA®

Overwhelmed by data?

Using and Protecting the "Right" Data

By: Andrew C. Lewis, CGFM, CPA, CIPP/G, PMP; Jeffrey C. Steinhoff, CGFM, CPA, CFE, CGMA; and Viral Chawda, MBA, MS, PgMP, PMP

Federal, state and local governments have access to incalculable amounts of data to manage programs; improve operations; fight fraud, waste and abuse; and better serve citizens. Whether data is generated in house or available externally, consider:

- Data is exploding. In 2014, it was estimated that 90 percent of the world's data had been produced within the previous two years.[1]
- As much as 80 percent of data is unstructured.[2] Today's ability to analyze unstructured data is a game changer.
- The federal government operates 224 core data centers, with an additional 10,000 non-core data centers.[3] Although efforts are underway to close approximately half of the non-core data centers, the amount of computing power and storage at the remaining centers is staggering.
- Data.gov, the federal government's open government data repository, includes more than 150,000 data sets available for public analysis.[4] In just the last two years since June 2015, the number of data sets has tripled.[5]
- The federal government is projected to spend more than $89 billion on information technology (IT) and IT-related projects in fiscal year 2017.[6]
- Cost of data storage is estimated at $25 per gigabyte of data.[7] This can be an appreciable, yet somewhat hidden cost to an organization.
- Governments and companies will spend more than $203 billion by 2020, compared to $130 billion in 2016, on managing, interpreting and analyzing available data.[8]

Figure 1. Data Lifecycle

Source: U.S. Department of Education's Privacy Technical Assistance Center

With the tremendous expansion of all types of data available to governments — from data on traffic patterns to information on key economic trends to national security intelligence — and a dramatically increased ability to use this data to manage government, it begs the question: Do we have, and are we protecting, the right data? The *financial* cost of over-protecting low-value, low-quality data can be high if all data is treated equally in terms of retention and safeguarding. Similarly, we've seen the high *national security* and *public-impact* cost of under-protecting highly sensitive data. This article addresses the importance of adopting effective and efficient data governance practices, addressing the reliability of protected data, and ensuring the proper balance of resources used to maintain and protect data that is not useful.

WHAT DATA ARE YOU PROTECTING?

Leading organizations have a handle on what data they have, its quality, and how it's being used and protected. They value reliable, relevant and timely data as a critical mission asset. Leading data governance actions include:

- **Standing-up a robust chief data officer function:** In leading organizations, the role of a chief data officer (CDO) is essential to governance by helping ensure data is properly managed and categorized.[9] Data management isn't just a function of the chief information officer (CIO), but permeates the entire organization. Governments are under-served if they appoint a CDO as an ancillary duty on the shoulders of already busy executive leaders, or if they place a CDO deep within the CIO organization. Ideally, the CDO is a full-time function, properly staffed with a healthy mix of data policy professionals and data scientists, and viewed as a peer to other C-level executives within the organization. In leading organizations, CDOs don't own data, its use or its security, but serve as catalysts and facilitators for good data and its protection. Everyone is critical to developing and protecting high-quality data. Those who own, use and rely on data serve on the front lines, supported by the CDO.
- **Establishing a data governance strategy:** Engage internal stakeholders, such as the agency head, senior political and career leadership, program managers and the inspector general, as well as external stakeholders, including service providers, contractors, legislators, citizen and public advocacy groups, other levels of government and the media. Establish the agency's overall commitment to sound data creation, reliability, availability, transparency, privacy, retention and protection. Embed this commitment in strategic planning and oversight.

  Support data governance through (1) clear roles and responsibilities; (2) well-designed policies and procedures; (3) fact-based trade-offs; (4) oversight and monitoring, including benchmarking; (5) education; (6) open communications; and (7) continuous reevaluation. In doing so, be highly cognizant of the potential for insider threats of data manipulation and improper disclosure, with the Edward Snowden and Chelsea Manning cases serving as poignant reminders.
- **Inventorying and cataloging data assets:** Leading organizations actively inventory all their data sets — both current and planned. They then catalog these data sets based on characteristics, such as reliability, sensitivity and use. It's critical to understand what data exists; but just as important, understand the risks that influence the levels of protection of that data.

  Data assets are often categorized based upon potential damage if the data is lost, stolen, compromised and/or released. Tiers for ranking potential damage may include (1) non-sensitive data,

available to the public; (2) non-sensitive data, for internal use only; (3) sensitive data, such as personally identifiable information or personal health information; and (4) highly protected data, such as data relating to national intelligence and security threats.

- **Managing data across its lifecycle:** A leading practice is to establish policies and procedures that recognize data through the different stages of its lifecycle, depicted in **Figure 1**.

Engagement between the CDO and data process owners, and robust data governance practices facilitate the ability to (1) capture relevant and useful data; (2) organize data into a structured data set; (3) utilize data in furthering agency missions and programs; (4) manage data in an archived status for research or comparison to trends in current data sets; and (5) identify data for destruction when it's no longer needed, and in accordance with records retention laws and policies.[10]

ARE YOU USING AND PROTECTING GOOD DATA?

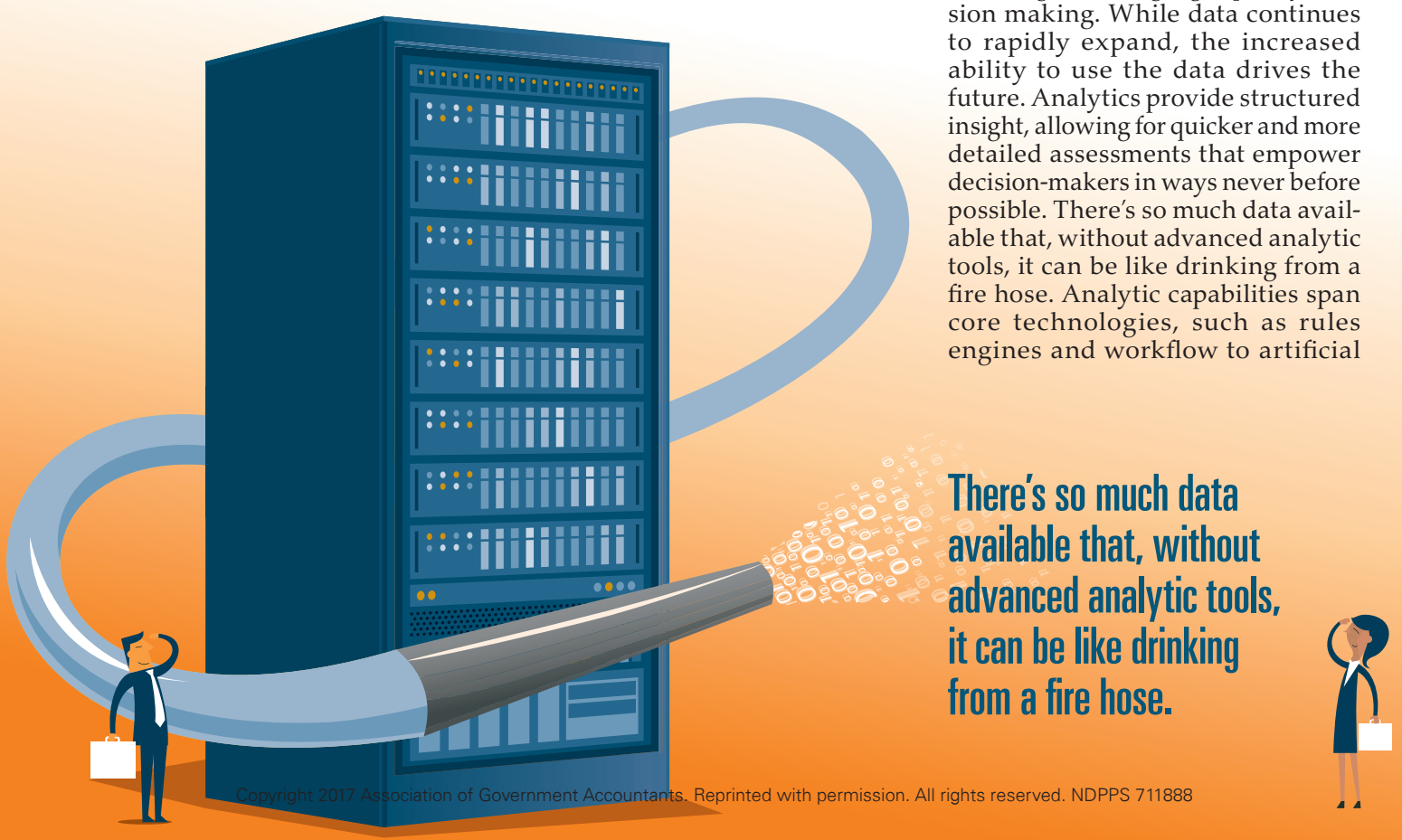
Leading organizations depend on data to oversee and improve programs, operate more efficiently and effectively, and provide insights on key trends and metrics. They recognize the importance of mastering the most fundamental element of any data and analytics initiative — capturing reliable, relevant and timely data. Poor underlying data can lead to misleading analytics, erroneous conclusions, and bad policy and decision-making.[11]

A 2015 *Governing* survey[12] of state officials and managers identified that data quality issues were widespread, with 69 percent of those responding to the question stating they "frequently" or "often" encountered data problems, and no one answering "rarely." The survey respondents identified these major causes of data quality problems:

- 17 percent: technology limitations
- 14 percent: management issues, such as apathy, and lack of supervision and accountability
- 12 percent: poor or lack of planning
- 11 percent: lack of training
- 10 percent: data entry issues, including a lack of data definitions and structure
- 9 percent: inadequate controls
- 7 percent: data-sharing and collaboration problems, including siloed systems

The survey found that most data problems are in program management systems rather than financial accounting systems. Among the program management agencies with data problems, social services and economic development ranked highest in the *Governing* survey. One state deputy controller observed that accountants are "trained in a discipline that places a high value on peer review, internal controls, [and] edits checks," but that this discipline is typically lacking among program administrators and staff outside of the chief financial officer (CFO) organization.[13]

Leading organizations turn data into insight, driving high-quality decision making. While data continues to rapidly expand, the increased ability to use the data drives the future. Analytics provide structured insight, allowing for quicker and more detailed assessments that empower decision-makers in ways never before possible. There's so much data available that, without advanced analytic tools, it can be like drinking from a fire hose. Analytic capabilities span core technologies, such as rules engines and workflow to artificial



There's so much data available that, without advanced analytic tools, it can be like drinking from a fire hose.

intelligence and machine learning supporting cognitive learning. This presents a spectrum of opportunities to process super data sets and apply evidenced-based learning in the context of predictive analytics and hypothesis generation.[14]

Analytic tools can do wonders, but the adage "garbage in, garbage out" still applies. Governments need to not only understand the quality of their own data, but the quality of external data being used as well. An example is using advanced analytic tools in the fight against improper payments, where data sets may be located across the federal government, as well as among state and local governments, contractors, and other stakeholders.[15]

Here are seven leading practices to help assess and improve data quality.

1. **Assess the data source:** Data generated from internal sources or from third parties pose different challenges in assessing data quality. Understanding the source of your data, and which internal or external key stakeholders to engage with, is necessary to best leverage data and make improvements where there are data shortfalls. Understand data flows and the controls in place to help ensure a high level of data quality.
2. **Communicate the importance of data quality:** In leading organizations, staff understand the importance of data quality and the pitfalls when data is not of a high quality. Identify the data that directly influences your analysis or conclusions. Communicate the importance of this information to internal and external stakeholders responsible for data input. To reiterate, the ability to capture and analyze unstructured data is a game changer. At the same time, the ability to capture vast amounts of non-standard data comes at a cost of ensuring data quality across a much broader landscape.
3. **Engage everyone in the fight for good data:** Train your front-line workforce and their supervisors how to spot bad data. Demonstrate the importance of good data through examples of how it helps improve government, such as positively impacting the citizen experience in interacting with government.
4. **Leverage technology to drive data quality:** Leading organizations leverage advanced analytics to identify and resolve data quality issues, especially focusing on data having the most impact. This becomes even more important as the volume, velocity and variety of data continues to grow. The amount of data is so enormous that advanced analytic tools provide the only realistic means of analysis. Similarly, the enormity also drives the use of these tools to help ensure data quality. Using built-in edit checks and selecting from structured lists of options, rather than allowing users to enter free-form text, are common first steps, but only scratch the surface. Enabling technology tools will only become more sophisticated as a means to drive data quality, and are a staple of leading organizations.
5. **Review data input processes:** Conduct periodic testing of data input processes by comparing data in the system to source records; again, leverage advanced analytic tools. Use the testing results as a key metric in individual or organizational performance evaluations of those responsible for maintaining the data. The ability to review data developed by organizations outside the government offers different challenges. Where possible, leading organizations work

Data-driven policy and decision making can increase the public's confidence in government, improve service delivery, and allow governments to tailor policies and programs to directly affected constituents or the most at-risk populations.

Figure 2. Four Dimensions of Assessing Data Usefulness

closely with these organizations to understand their processes to address data quality and share lessons learned. When this isn't possible, they look for secondary sources that may address data quality, such as outside reviews and evaluation studies.

6. **Link data to other data sets:** Improve data interoperability by linking data between different systems. Data errors are detected more readily when compared to a complementary data set. Find ways to break down data silos by linking data and by investigating gaps.
7. **Engage a third party:** If your organization doesn't have in-house capability, use a third party steeped in data assessment and application of advanced data analytic tools, such as the inspector general, or independent public accounting and management consulting firms, to help evaluate the quality and reliability of key data. The goal is to have insight into the strengths and weaknesses of data quality and reliability processes, leading to high-quality data.

ARE YOU PROTECTING DATA THAT'S VALUABLE?

Not all data is created equally. Although data can be plentiful in any organization, establishing what data is useful to your mission and decision-making processes is an important factor in deciding how much to spend in developing, maintaining, storing and protecting specific information. Organizations may find they have too much data, while lacking key information.

Useful data can profoundly impact the way government operates. Data-driven policy and decision making can increase the public's confidence in government, improve service delivery, and allow governments to tailor policies and programs to directly affected constituents or the most at-risk populations. Think of it as adding greater value by (1) providing predictive intelligence and trend analysis by looking to the future and seeing around corners; (2) pinpointing anomalies to help prevent and detect fraud, waste and abuse; (3) targeting areas of highest value or risk for increased scrutiny; (4) staying in front of issues as they evolve; and (5) shifting the audit paradigm by having management use common audit tools, such as data mining, to identify and address anomalies because, after all, management and not the auditors own the information.[16]

Leading organizations assess data usefulness along four dimensions, as depicted in **Figure 2**.

CLOSING THOUGHTS

Like any asset, data must be created, managed and maintained. It also needs to be reliable, relevant and timely, to be useful. There's so much data today that leading organizations keep their finger on the pulse of what's available, how it can be used and the available enabling analytic tools. Governments face fiscal[17] and service delivery challenges for which the ability to fully leverage data can be a game changer.

Effective data governance practices — including continually evaluating the reliability and usefulness of data — can increase the odds that the right data is being maintained, used and protected, and that resources are not wasted on irrelevant or simply bad data. Also, through advanced data analytics, leading government organizations can turn data into powerful insights to manage programs and operations, and enhance the citizen experience, while helping ensure data quality and protection.

The 2012 World Economic Forum[18] declared data to be an economic asset like currency or gold. Come to think about it, perhaps this has always been the case. In the words of founding father President Thomas Jefferson: "Information is the currency of democracy." So, treat data like currency or gold to serve the American public and our nation's democratic ideals.

Endnotes

1. "Big data, for better or worse: 90% of the world's data generated over the past two years," *Science Daily*, May 22, 2013.
2. "The biggest data challenges that you might not even know you have," by Christie Schneider, *IBM Cognitive Enterprise*, May 26, 2016.
3. "Data Center Consolidation: Agencies Making Progress, but Planned Savings Goals Need to Be Established," U.S. Government Accountability Office (GAO), GAO-16-323, March 3, 2016.
4. www.data.gov/metrics; accessed July 31, 2017.
5. www.data.gov/metric/federalagency/dataset-published-per-month; accessed Aug. 8, 2017.
6. "Information Technology: Improved Implementation of Reform Law Is Critical to Better Manage Acquisition and Operations," GAO-17-263T, Dec. 6, 2016.
7. "The Cost of Managing Unstructured Data," Enterprise Strategy Group, by Kerrey Dolan and Brian Garrett, May 2014.
8. International Data Corporation (IDC) *Worldwide Semiannual Big Data and Analytics Spending Guide*, 2015.
9. "The Chief Data Officer's Path to Success," by Viral Chadwa, Governing.com, Aug. 19, 2016.
10. "Best Practices for Data Destruction," U.S. Department of Education Privacy Technical Assistance Center, PTAC-IB-5, May 2014.
11. See Endnote 9.
12. "The Causes, Costs and Consequences of Bad Government Data," by Katherine Barrett and Richard Greene, Governing.com, June 24, 2015.
13. Ibid.
14. "Embracing Game Changers," by Jeffrey C. Steinhoff, Kirke E. Everson, Viral Chawda, and Joseph M. Ward, *Journal of Government Financial Management* (*Journal*), spring 2017.
15. "Are You Combat Ready to Win the War Against Improper Payments?," by Danny Werfel and Jeffrey C. Steinhoff, *Journal*, summer 2014.
16. "Calling All Government Financial Managers to a More Analytic Role as Highly-Valued Business Advisors!," by David A. Fitz, James P. Hauer III, and Jeffrey C. Steinhoff, *Journal*, summer 2015.
17. For example, "The Nation's Fiscal Health: Action is Needed to Address the Federal Government's Fiscal Future," GAO-17-237SP, Jan. 17, 2017.
18. "Big Data, Big Impact: New Possibilities for International Development," World Economic Forum, 2012.

Andrew C. Lewis, CGFM, CPA, CIPP/G, PMP, a member of AGA's National Executive Committee and Montgomery/P.G. County Chapter, is a partner in KPMG LLP's federal practice. He is an executive fellow of the KPMG Government Institute and an adjunct professor at The George Washington University.

Jeffrey C. Steinhoff, CGFM, CPA, CFE, CGMA, an AGA Past National President and member of AGA's Northern Virginia and Washington DC chapters, is managing director of the KPMG Government Institute. During a 40-year federal career, he was assistant comptroller general of the U.S. for Accounting and Information Management, led GAO's largest audit unit, had responsibility for developing government auditing and internal control standards, and was a principal architect of the CFO Act. He founded AGA's CGFM program and received the Robert W. King Memorial Award, AGA's highest honor. He is an elected NAPA fellow.

Viral Chawda, MBA, MS, PgMP, PMP is a managing director in KPMG's Data & Analytics Center of Excellence. For more than 19 years, he has helped government and commercial organizations with optimization and monetization of data assets across a broad range of applications, including finance, risk, operations, portfolio management and innovation. He is also an executive fellow of the KPMG Government Institute.

About the KPMG Government Institute

The KPMG Government Institute was established to serve as a strategic resource for government at all levels, and also for higher education and not-for-profit entities seeking to achieve high standards for accountability, transparency, and performance. The Institute is a forum for ideas, a place to share leading practices, and a source of thought leadership to help governments address difficult challenges such as performance management, regulatory compliance, and fully leveraging technology.

kpmg.com/us/governmentinstitute

Contact us

Andrew C. Lewis
Partner
KPMG LLP
T: +1 202 533 4886
E: aclewis@kpmg.com

Jeffrey C. Steinhoff
Managing Director
KPMG LLP
T: +1 703 286 8710
E: jsteinhoff@kpmg.com

Viral Chawda
Principal
KPMG LLP
T: +1 214 840 2000
E: vchawda@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates.

kpmg.com/socialmedia