



# Using Data

Mark A. Abramson

### *Highlights*

- The use of data has risen exponentially. However, government agencies face challenges in transforming data into actionable insights.
- With the increased use of data, the challenges of handling data have also increased. As government makes open data more accessible, challenges include finding data experts and managing data accessibility, data quality, and data sharing.
- Data sharing by the private sector, data sharing among government agencies, and the government's capacity to manage and analyze its increasing volumes of data will be critical factors in the years ahead.

# USING DATA

By Mark A. Abramson

*In 2016, the Department of Justice (DOJ) and the Department of Health and Human Services (HHS) worked together to uncover a money laundering and health care fraud case involving a \$1 billion scheme being perpetrated by a ring of Miami-based health care providers. Caryl Brzymiakiewicz, chief data officer in the Office of the Inspector General (OIG) at HHS, said, “Three co-conspirators had figured out how to try to hide in the data. But it was really about using data analytics and partnering with DOJ and the FBI to uncover the money laundering, to understand in the data what was happening, really understand between the provider and all of the networks really what was going on.”<sup>1</sup>*

*This effort involved working closely with Centers for Medicare & Medicaid Services (CMS), which gave their claims to the HHS OIG who then used the claims in their investigation with the Department of Justice. Jessica Kahn, director of the Data and System Group at CMS, said that it is not always easy to share data across agencies. “We put it (our data) in the cloud because I want people to use it,” said Kahn.<sup>2</sup> She said she gives her data to Brzymiakiewicz because she wants the Inspector General “to catch the bad guys.”*

## INTRODUCTION

### Data as a Strategic Asset

The chapters in this book are closely interrelated. As seen in the previous chapter, the rapid movement to “going digital” over the last twenty years served as a key enabler to the increased capability of government to collect and analyze data. New technologies, also discussed in Chapter Two, dramatically reduced the cost of collecting and reporting data. The ability of government to collect and analyze data has similarly been a valuable tool enabling the performance management movement to shift the emphasis from complying with reporting requirements to generating more useful data that informs performance improvement efforts. Performance management is examined in Chapter Four.

This chapter focuses on data collected and used by government managers and decision makers in managing their organizations. In 2018, the President’s Management Agenda designated “Leveraging Data as a Strategic Asset” as a Cross-Agency Priority (CAP) Goal. In its description of the CAP

Goal, the Administration set out three key opportunities to more effectively use data in coming years:<sup>3</sup>

- Develop a long-term enterprise Federal Data Strategy to better govern and leverage the federal government's data.
- Enable government data to be accessible and useful for the American public, businesses, and researchers.
- Improve the use of data for decision making and accountability for the Federal Government, including policy making, innovation, oversight, and learning.

While all three are interrelated, and the development of a Federal Data Strategy will have an influence on the use of data by decision-makers, this chapter focuses on what the CAP Goal describes as “providing high quality and timely information to inform evidence-based decision-making and learning.” This chapter will not address the host of policy and legal questions, such as citizen privacy and the security of data, collected by statistical agencies, and administrative data collected by other agencies (privacy and security are addressed in Chapter Two and Chapter Seven).

In order to understand the evolution of data and data policy, it is important to note that the root of federal data policy goes back to the Paperwork Reduction Act of 1980 and Office of Management and Budget (OMB) Circular A-130, *Management of Federal Information Resources*. Both the Paperwork Reduction Act and the OMB Circular established the following key principles:

- sound information management policies are crucial
- government will provide free and open access to data
- data will be treated as a strategic asset by government

The key challenge now is transforming data into actionable insights for government executives. In short, how can government make sense of its vast and growing amounts of data to develop new understandings that inform decisions? While new technologies now allow for the collection, analysis, and sharing of real time data, agencies face the challenge of making data relevant and meaningful to decision makers.

The last two decades have been characterized by a more robust supply of useful data and performance information that can serve as a foundation for more evidence-based insights and decisions in the future. Government policy in recent years has encouraged the greater availability of open data, which has contributed to the growing supply of useful information. The increased emphasis on open data occurred via administrative and legal channels, including policies surrounding Open Data commitments, the adoption of the Digital Accountability and Transparency Act (DATA Act), and additional commitments to make routine administrative data more widely available via channels such as Data.gov.

## Organization of Chapter

This chapter addresses major developments in how government has evolved its use of data between 1998 to 2018. As seen in the chart, “Evolution of Data: 1998–2018,” the evolution of data can be divided into three phases:

- **Early action:** This phase was characterized by an important shift from simply collecting and reporting data to using and analyzing data. Government organizations at the federal, state, and local levels all demonstrated an increased interest in timelier, more useful data. This emphasis was seen in the creation of PerformanceStat initiatives in localities across the nation. During this phase, the federal government also continued its interest in the use of data generated by state and local governments.
- **Expansion:** Based on the increased production of data, government organizations began to focus on new ways to more effectively use the datasets that were being produced. New, more effective uses of data included increased use of analytics, data visualization tools, and big data.
- **Institutionalization:** Based on government’s increased experience with the creation and use of data, government policies needed to change. These changes resulted in a series of new policies, increased use of open datasets, and the creation of chief data officer positions.

### Evolution of Data: 1998—2018

1998



#### **Early Action: The Shift to Analyzing Data**

---

- A Shift from Collecting and Reporting Data to Using and Analyzing Data
- Increased Interest in Timelier, More Useful Outcome Data
- Increased Use of State Government Data by Federal Government

2005

2005



#### **Expansion: Toward More Effectively Analyzing Data**

---

- Increased Use of Analytics
- Increased Use of Visualization
- Increased Use of Big Data

2010

2010



#### **Institutionalization: Making Open Data More Accessible**

---

- New Policy Guidance and Laws
- Increased Use of Open Datasets by Government
- Creation of Data-Focused Governance Positions

2018

## EARLY ACTION: THE SHIFT TO ANALYZING DATA

While interest in performance management has a long history (as described in Chapter Four), great strides were made in the early 1990s to mandate and stimulate the collection of data which could be used in performance management systems. A notable impetus to organized data collection (for use in performance management) was the passage of the Government Performance and Results Act of 1993 (GPRA). GPRA mandated the development of agency strategic plans, annual operating plans, performance measures, and reporting systems.

### **A Shift from Collecting and Reporting Data to Using and Analyzing Data**

The late 1990s and early 2000s witnessed the shift in emphasis from the collecting and reporting of data to the analysis and use of data. These uses included identifying and understanding problems in need of attention in specific government activities and setting goals to measure progress. Traditionally, government organizations have collected administrative data on operations, but the use of such data was often limited. This began to change in the 1990s.

A prime example of the movement to using and analyzing data is the implementation of PerformanceStat initiatives that were created at the local level and spread throughout the nation in the 1990s. Robert Behn characterizes PerformanceStat as action-oriented, data-informed problem solving meetings in government agencies, which focus on using data to find problems in need of attention.<sup>4</sup> Behn notes that one of the key components of these Stat systems is the use of data to analyze specific aspects of an organization's performance.

The first well-known and most widely publicized "Stat" initiative was CompStat, which was created by the New York City Police Department in the mid-1990s. Creation of "Stat" initiatives followed in other localities throughout the next decade. Numerous IBM Center reports chronicled the development of these Stat systems, also discussed in Chapter Four, including a 2001 examination of CompStat by Paul O'Connell, *Using Performance Data for Accountability: The New York City Police Department's CompStat Model of Police Management*.<sup>5</sup>

O'Connell noted that a fundamental, essential principle of CompStat was the collection of accurate and timely information (data), and the meaningful analysis and dissemination of the data. It is hard today to recall the state of technology in the mid-1990s and early 2000s. One of the consistent recommendations from the early years of the Stat movement was the need to have an updated information technology infrastructure in place for a Stat system to

operate effectively. Equally important, according to O'Connell, was the need to compile timely and accurate data to share in advance of the Stat meeting (meetings held among senior staff and front-line managers to discuss the data presented and take appropriate action). The PerformanceStat movement also represented the start of increased attention to the need to make data transparent and publicly available.

The second well-known PerformanceStat program was CitiStat, which originated in Baltimore and was spearheaded by former Mayor Martin O'Malley. In his 2003 report, *The Baltimore Citi-State Program: Performance and Accountability*, Lenneal Henderson also focused on the importance of data collection to the Stat process.<sup>6</sup> Henderson wrote that a key to the success of the Stat model is the identification, collection, and analysis of agency performance and personnel data. While it is easy to underestimate the challenges facing the early Stat programs, one of CitiStat's accomplishments was the creation of a computerized information network to collect biweekly data from agencies. This biweekly data generated analyses of performance trends used in early Geographical Information System (GIS) formats to examine the distribution of city services, needs, and challenges. Henderson recommended that CitiStat data needed to be better compiled and simplified for both internal use and broader public use. He also observed that the next challenge for CitiStat data would be to develop indicators to mark progress on citywide mayoral initiatives such as crime reduction, public safety and security, housing, and health care. The first wave of Stat programs tended to focus on the progress and accomplishments of single departments.

The Stat movement has also been used in school systems. In their 2007 report, *The Philadelphia SchoolStat Model*, Christopher Patusky, Leigh Botwinik, and Mary Shelley examined the Philadelphia School Stat program.<sup>7</sup> This Stat program identified key performance indicators that quantified school and student performance in multiple areas. The report's authors found that the collection of data did impact the culture of the organization. The authors wrote that, based on their interviews with school district staff, "the District's culture has begun to operationalize the regular use of data as part of its management routines, and this represents an important step toward establishing a data-driven organizational culture."<sup>8</sup> Data began to be reviewed by both staff and students. SchoolStat and other Stat programs were challenged by the faulty assumption that staff would make effective use of data generated by the initiative.

## **Increased Interest in Timelier, More Useful Outcome Data by Federal Government Agencies**

During this same time period, the early and mid-2000s, federal agencies also increased their emphasis on outcome-focused management, in response to GPRA as well as changes in authorizing legislation. In a 2004 report, *How Federal Programs Use Outcome Information: Opportunities for Federal*

*Managers*, Harry Hatry, Elaine Morley, Shelli Rossman, and Joseph Wholey examined how federal programs used outcome information.<sup>9</sup> The authors found that the quality of data continued to be a major problem in successfully assessing program outcomes. Specifically, Hatry and his colleagues found:

- datasets were often “old” by the time they reached program managers
- even if the datasets were not “old,” the timing as to when the data became available for use by program managers was an issue
- some of the datasets were not actionable to be useful to many program managers

These findings led Hatry and his colleagues to recommend that timelier data be sought, and that the data be presented in a user-friendly form. The federal government would spend the remainder of the decade addressing the accuracy and timely collection of data.

## **Increased Use of State Government Data by Federal Government**

Prior to cities implementing Stat initiatives, the federal government had begun developing strategies for better using information from state governments. In a 2003 report, *Strategies for Using State Information: Measuring and Improving Program Performance*, Shelley Metzenbaum examined the use of state information by the federal government.<sup>10</sup> Metzenbaum found that state performance information was helpful to federal and state government when used to identify successes and problems, as well as to trigger focused follow-up inquiries that enabled everyone in the delivery systems—from the front-line to the federal program office—to better understand the causes of problems and contributors to success.

## **EXPANSION: TOWARD MORE EFFECTIVELY ANALYZING DATA**

During this time period, substantial progress was made on three fronts. First, great strides were made in better analyzing the data collected. Second, improvements emerged in the use of visualization tools to better communicate the data which had been collected and analyzed. And third, the use of big data expanded the capability of government to access and analyze large datasets with increased speed.

## Increased Use of Analytics

The mid-2000s saw significant breakthroughs in data-capturing technologies, data standards, and data storage, accompanied by improvements in modeling and optimization science. With the increase in available data, the challenge became placing the data in context to understand its implications for decision-making. As a result, noted Tom Davenport and D.J. Patil, new opportunities arose for the use of analytics. In their 2008 report, *Strategic Use of Analytics in Government*, Davenport and Patil defined analytics as “the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions.”<sup>11</sup>

The increased interest in the use of analytics during this time period reflected the realization that many government agencies had considerable administrative data at their disposal. *Administrative data* refers to information collected primarily for administrative (not research or statistical) purposes collected by government organizations as part of their transactional activities and record keeping. Examples of administrative data include information gathered from tax filings, registrations, and in connection to applications for government benefits and other administrative activities. However, most agencies did not analyze administrative data in detail, which would have enabled them to identify opportunities to improve services or increase revenue.

A major problem identified during this time period, which continues to this day, is the limited availability of skilled resources in government agencies to analyze data. A key recommendation from Davenport and Patil, echoed by other reports discussed in this chapter, was that “...government organizations need to develop a cadre of analysts—both professional and amateur.”<sup>12</sup>

During this same time period, the IBM Center and the Partnership for Public Service undertook a multi-year project examining the use of data and analytics by government agencies. A key lesson that emerged from these studies was the insufficiency of merely collecting and reporting data. Government had indeed improved its ability to collect and store data. In their 2011 report, *From Data to Decision: Power of Analytics*, the Partnership wrote, “... we learned that data is only the starting point. The data need to be analyzed, turned into information and made accessible to staff and executives, and the data also is needed to meet varying needs and to be understandable to different audiences.”<sup>13</sup>

The Partnership conducted a series of interviews with federal executives for this 2011 report. The interviews identified several significant challenges regarding data for use in analytics:

- ownership of data
- availability of data
- maintaining data integrity

The 2012 study, *From Data to Decisions II: Building an Analytics Culture*, from the Partnership and the IBM Center echoed the recommendation from Davenport and Jarvenpaa—skilled staff were a critical piece in the effective use of analytics in government.<sup>14</sup> This report found that the government was increasing its use of analytics to document what it does, assess effectiveness, and determine measurement processes. Based on the use of analytics, agencies were identifying changes which needed to be made to improve program performance and achieve better results.

The need to “embed” an analytics culture into government was a key focus of the 2013 study *From Data to Decisions III*.<sup>15</sup> This report emphasized encouraging the use of data by employees. The report recommended that employees be able to easily see, combine, analyze, and use data. The report found that, “Leaders and managers should demand and use data and provide employees with targeted on-the-job training.”<sup>16</sup>

A good example of the use of analytics in the public sector was the increased interest in predictive policing. Predictive policing can be viewed as a descendent of CompStat in which crime data was used as one input into the deployment of police officers in the field. In her 2013 report, *Predictive Policing: Preventing Crime with Data and Analytics*, Jennifer Bachner wrote that the fundamental notion of predictive policing “is that we can make probabilistic inferences about future criminal activity based on existing data.”<sup>17</sup> Bachner found that predictive policing faced several major challenges which focused on the quality of data and the training of analysts to use the data. Bachner found that collecting and managing large volumes of accurate data pointed to one major challenge facing the use of predictive policing. A second major challenge was ensuring that analysts possessed sufficient domain knowledge about law enforcement to analyze the available data. An additional challenge emerged as maintaining adequate analytical resources to use the data, a reoccurring theme in studies on the effective use of data. In her recommendations, Bachner emphasized the need to collect accurate and timely data, and to designate leaders who were committed to the use of analytics.

## Increased Use of Visualization

A major factor in increasing the use of data by government in this time period was making it more accessible for users, enabled by the significant advance in the visualization of data. While data visualization has a long history, advances in both hardware and software made it substantially easier to use.

One type of visualization is enabled by the increased use of GIS and, more specifically, geo-coding of data. In a 2010 report, *Using Geographic Information Systems to Increase Citizen Engagement*, Sukumar Ganapati examined the use of GIS.<sup>18</sup> The ability of citizens to visually see transit routes, obtain transit information, and to provide citizen-volunteered information was made possible by government agencies providing data in standardized formats. This

access to public domain data enabled government agencies and third parties to develop GIS apps aimed at citizens. The importance of open data and standardized formats is discussed further, later in this chapter.

Another type of data visualization is the use of dashboards by government managers to track performance. In a 2011 report, *Use of Dashboards in Government*, Sukumar Ganapati assessed the impact of dashboards.<sup>19</sup> Ganapati writes, “Organizational dashboards are often likened to dashboards in plane cockpits and cars, which allow the pilot or the driver to see instant information about various metrics...and make travel adjustments or spot vehicular issues on the fly.”<sup>20</sup> He found that the quality of data was key to the credibility of dashboard performance measures. Like other research on the use of data, he noted that dashboards were only tools and their effectiveness depended on the use by managers.

In her 2013 report, *The Use of Data Visualization in Government*, Genie Stowers noted that effective data visualization, or graphic display, has been used to understand data patterns since 1854 when a doctor in London mapped cases of cholera.<sup>21</sup> Her report tracked the movement toward increased use of visualization in government. She wrote, “The movement is the result of numerous converging trends—the open data and transparency movements, growing citizen engagement with data, new tools for data mining and analysis that use ever larger datasets, advances in web graphic technology and interactive online mapping and graphing, and new awareness of the need for more proactive citizen engagement.”<sup>22</sup>

## Increased Use of Big Data

The mid-2010s saw an increase in the use of big data. As noted earlier, technological advances made it dramatically easier to collect and store data, with the cost of storing data falling sharply over the years. In his 2014 report, *Realizing the Promise of Big Data: Implementing Big Data Projects*, Kevin Desouza reported that not only were storage devices cheaper, significant advancements in the science of databases and information retrieval emerged as well.<sup>23</sup>

In his report, Desouza defined big data as an evolving concept that refers to the growth, value, and speed of data, and how data can be analyzed to optimize business processes, create customer value, and mitigate risks. Desouza quotes authors Viktor Mayer-Schonberger and Kenneth Cukier that “big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of values, in ways that change markets, organizations, the relationships between citizens and governments, and more.”<sup>24</sup>

In their 2016 report, *Ten Actions to Implement Big Data Initiatives: A Study of 65 Cities*, Alfred Ho and Bo McCall referred to big data as “using massive amount of data to conduct analyses so that the data patterns and relationships can be used for classification, clustering, anomaly detection,

prediction, and other analytic needs in decision making.”<sup>25</sup> Ho and McCall also reported that, with the advancement of computing technologies and the emergence of many data analytic tools, user-friendly platforms can be used to conduct more sophisticated program and customer analysis.

Ho and McCall surveyed 65 cities to understand their use of big data and analytics. They found that 75 percent of the cities surveyed reported having undertaken big data initiatives, including increased use of analytics, better integration of data with budgeting, and using a team approach or multi-departmental governance structures for their data initiatives. Their survey also found that many cities were creating chief data officer positions to lead these data initiatives. Cities were also increasingly providing citizen-friendly ways to visualize city and access data, as well as empowering citizens to conduct their own data inquiries and analysis of city-generated data.

While Ho and McCall found that big data was being used in the cities they surveyed, and had much potential, a variety of issues involving data began to surface. The increase in the availability of data created new ethical and legal challenges in both the public and private sectors. These issues included potential privacy and individual rights infringement, hidden inequity and discrimination in algorithm-driven decision making, and potential conflicts between efficiency, customization, and equal access to government services by all. Specific privacy issues include how data should be collected, stored, and analyzed, as well as how data should be shared with non-government entities.

### **Case Study in Collecting and Using Data: The American Recovery and Reinvestment Act of 2009**

A key event in advancing government's ability to collect and use federal financial and performance data was the implementation of the American Recovery and Reinvestment Act of 2009 (Recovery Act). An implementation goal set by the Office of Management and Budget was that the use of all Recovery Act funds be transparent to the public and that public benefits of the Act be reported clearly, accurately, and in a timely manner. Oversight of the implementation of the Act was assigned to the newly created independent Recovery Accountability and Transparency Board (Recovery Board), comprised of agency Inspectors General. The Recovery Board had the responsibility to establish and maintain a user-friendly, public-facing website, Recovery.gov, to foster accountability and transparency in the use of Recovery Act funds over the course over the six-year initiative. The Recovery Board also created a Recovery Operations Center which was responsible for cross-referencing data from recipient reports and other government databases to detect fraud and misuse of funds.<sup>26</sup>

The experience of implementing the Recovery Act was a significant “learning experience” for the federal government and helped lead to the DATA Act of

2014 that codified many of the lessons learned by the Recovery Board. In addition, implementation of the Recovery Act demonstrated many of the capabilities and tools discussed in this chapter:

- **Standardization of data collected**

The Recovery Board required the recipients of its funds to input 99 fields of numerical and narrative data related to six dimensions of spending. By generating detailed, multilayered recipient reports tracking Recovery Act financial data, the Act acted as “proof of concept” for future, more ambitious public transparency initiatives regarding federal spending.

- **Use of predictive analytics**

The Recovery Operations Center used a variety of tools to mine more than 25 government and open-source databases, looking for anomalies and indicators of fraud or waste.

- **Use of new technologies**

According to Earl Devaney, former chair of the Recovery Board, the success of the Recovery Operations Center was based on the Board’s ability to find the right set of tools to collect, manage, and analyze numerous datasets.

- **Mapping**

The Recovery.gov website provided comprehensive geospatial capability for citizens to find Recovery Act spending in their localities and for use by the Recovery Board to map incidents of fraud and waste.

- **Cloud computing**

The Board had a clear need to seek new levels of efficiency and cost savings in the collection and analysis of data. Their efforts were an early demonstration of the value of cloud computing. The move to the cloud meant that the Board no longer had to manage the Recovery.gov’s physical data center and related computer equipment.

- **Continuous monitoring**

The concept of continuous monitoring helped reduce the reliance on human analysts to perform predictive analytics. By leveraging big data systems, the continuous monitoring process eliminated the need for additional interpretation of data before taking action.

## Identification of Challenges in the Use of Data

A 2018 study by the Pew Charitable Trusts, *How States Use Data to Inform Decisions*, reported the challenges that government executives at the state level faced in using data, were similar to those faced at both the federal and local levels.<sup>27</sup> The report found the following major challenges:

- **Challenge One: Staffing.** Few state employees were experienced in both policy and data analytics. Many states reported that existing staff lacked skills in data analytics or the ability to interpret data findings to make policy recommendations.
- **Challenge Two: Data accessibility.** Many state agencies had archaic data systems, some developed in the 1980s, which made it very difficult to access and use data.
- **Challenge Three: Data quality.** Data quality issues impaired the analyses of data. Many state databases suffered from quality issues which made them difficult to use and interpret.
- **Challenge Four: Data sharing.** If a state agency wanted to make quality data accessible, a combination of problems including organizational culture, laws, or other factors often prohibited the data from being shared.

While the Pew report found that data were indeed being used in strategic ways in state decision making, the above four challenges all need to be addressed to enhance more effective use of data. The report contains a series of recommendations, including the need for a more organized and centralized approach to data in the future. Key actions for state leaders include the development of “governance structures to guide data use and access while also prioritizing privacy” and the need to “take stock of their data systems and perform an inventory of data sets.”<sup>28</sup> A major part of the data governance process involves the need to ensure that quality data could be accessed and used by stakeholders. Key steps include improving data quality and accessibility, developing an enterprise view of data, and establishing data sharing agreements.

As has been seen throughout this chapter, concerns continue about the need to build government’s capacity to effectively use data. The Pew report recommended hiring new staff skilled in data analytics. The report also recommended that funding be dedicated to support data-driven projects. The federal government now faces the same challenges of skilled staff and adequate funding for data-driven projects.

In addition to addressing governance, staff capacity, and the quality of data, the Pew report made a series of recommendations concerning the use of data. While this chapter focuses on the availability (or supply) of data, Chapter Four addresses issues surrounding the use (demand) of data. The Pew report recommended an increase in the use of visualization techniques in charts, dashboards, and reports to make the data easier for decision makers to analyze and understand. Findings from the analysis of data can inform, guide, or alter decisions. The Pew report concluded with recommendations that agencies should create an organizational culture that prioritizes data collection, and that new legislation and policies are needed to support data use.

## INSTITUTIONALIZATION: MAKING OPEN DATA MORE ACCESSIBLE

Institutionalization is reflected in several major developments in recent years: a series of directives and new policies at the federal level to open up government datasets, increased use of existing administrative datasets by government executives, and the creation of chief data officer positions.

### New Policy Guidance and Laws

Starting in 2009, a series of federal policies and laws contributed to the opening of data sources to the public. In addition to becoming available for use by the public, these open datasets also proved highly useful to government agencies as they delivered their missions. In addition to improving accessibility to these datasets, emphasis was also placed on increasing the quality of data. Key policy directives included:

- **Open Government Directive:** This 2009 directive required agencies to publish more information online in open and accessible ways. It also required agencies to increase the amount of high-value datasets available to researchers and directed OMB officials to create an interagency process for sharing and coordinating data policies.<sup>29</sup>
- **Open Data Policy—Managing Information as an Asset:** This 2013 directive was to promote interoperability, accessibility, and openness in regard to data. Agencies were required to use data standards and extensible metadata for information creation and collection efforts, and to ensure information stewardship.<sup>30</sup>
- **Guidance for Providing and Using Administrative Data for Statistical Purposes:** This 2014 directive called for greater collaboration between program and statistical offices and encouraged agencies to promote the use of administrative data for statistical purposes.<sup>31</sup>
- **The Digital Accountability and Transparency Act of 2014 (DATA Act):** This law aims to make federal spending information more accessible and transparent. The law requires the Department of the Treasury to establish common standards for financial data provided by all government agencies and to expand the amount of data that agencies must provide to USASpending.gov, which is discussed below.<sup>32</sup>

### Increased Use of Open Data Datasets by Government

The policy directives and new laws outlined above have had a significant impact on making government datasets more widely accessible to the public and more user-focused. These datasets also proved useful to some government agencies in achieving their missions. Two types of new web portals—

data repository websites operated by the federal government, and external data repository websites which used government data—saw increased use.

Key government-hosted data web portals include:

- **USASpending.gov**, initially launched in 2007 in response to the Federal Funding Accountability and Transparency Act of 2005 (FFATA) mandated that federal contract, grant, loan, and other financial assistance awards of more than \$25,000 be displayed on a publicly accessible and searchable website to give the public access to information on how its tax dollars were spent. The Digital Accountability and Transparency Act of 2014 expanded FFATA by establishing government-wide data standards for financial data and providing consistent, reliable, searchable, and accurate data. The website was relaunched in April 2018 with expanded analytical tools and visualization capabilities.
- **Data.gov**, launched in 2009 by the Obama Administration to improve public access to high-value datasets generated by the federal government. In response to the 2013 Federal Open Data Policy discussed above, all future government data must be made available in open, machine-readable formats, while continuing to ensure privacy and security.
- **HealthData.gov**, created in 2012 as an outgrowth of the Health Data Initiative (HDI) established within the Department of Health of Human Services to make health data more available. At its launch, the website contained over 2,000 datasets. The website makes high-value health data more accessible to entrepreneurs, researchers, and policy makers in the hopes of better health outcomes for all.

In addition to the establishment of government data web portals, there has been an increase in the number of commercial web portals using government data. These new websites benefited greatly from the Open Data policies of the late 2000s and early 2010s, which made government datasets more available and accessible. Notable non-government data portals include:

- **DataUSA**, created in 2014 as a comprehensive website and visualization engine for publicly available U.S. government data. The site provides an easy-to-use platform that allows individuals to conduct their own analyses and turn data into knowledge.
- **USAFacts**, created in 2018 as a non-partisan, not-for-profit civic initiative which presents a data-driven portrait of the American population, government's finances, and government's impact.

## Creation of Data-Focused Governance Positions

In her 2018 report, *Data-Driven Government: The Role of Chief Data Officers*, Jane Wiseman found that there are currently few individuals in the federal government with the official title of chief data officer (CDO) at the departmental level.<sup>33</sup> There is, however, a clear trend to designate individuals who will have data responsibilities, some with the CDO title. While few of

these individuals will have the same set of responsibilities, most CDO-type positions will have a portfolio of activities that include data governance, data analytics, geographic information systems, data culture, smart technology, data infrastructure, and digital services. There has also been an increase in the number of data scientist positions throughout the federal government.

## LESSONS LEARNED

Based on our review of research on this topic over the past twenty years, we identified two key lessons:

**First, data made available for public use has also proved to be useful to government organizations themselves.** A case study on the impact of the transparency requirements from the American Recovery and Reinvestment Act of 2009 found that government officials became the primary users of the Recovery Act data because it allowed them to manage and track federal spending in near-real time. In a 2012 report, *Recovery Act Transparency: Learning from States' Experience*, Francisca Rojas found that spending transparency became institutionalized in some states and at the federal level in response to reporting requirements and that the data was used effectively by government executives.<sup>34</sup>

**Second, standardization of data provides a crucial step in the collection and sharing of data.** Significant strides have been made since the implementation of the Recovery Act. In describing his experience as chair of the Recovery Accountability and Transparency Board to monitor Recovery Act spending, Earl Devaney noted the difficulty of harmonizing spending data across agencies with different data standards. He concluded that, in order to effectively track money and to use data to make better-informed decisions, government will have to reevaluate how its databases interact and leverage each other. Many of the lessons learned in implementing the Recovery Act influenced the DATA Act of 2014, which moved government to a more cohesive, centralized accountability framework to track and oversee spending with standardized data formats.

## LOOKING FORWARD

A variety of important issues appear on the horizon regarding the future use of data by government agencies.

How can government use data collected by the private sector? To date, the emphasis has been on making data “open” from the government to the public, including the private sector. A future challenge will face the private sector to make its data “open” to the government and other users. This shar-

ing would create the possibility of effectively combining data collected by the government and the private sector.

A series of issues relate to sharing of data between federal government agencies themselves, between the federal government and other levels of government, and between local governments. Presently, the sharing of data between federal agencies poses problems because of statutory limits on sharing data. Proposed legislation, the Foundations for Evidence-Based Policymaking Act of 2017, would ease barriers which currently make the sharing of data between agencies difficult.<sup>35</sup>

The capacity of the federal government to both manage and analyze its data continues to be a major issue, as discussed earlier in findings from the Pew Charitable Trusts report on the state use of data. Another report, the 2017 report of the Commission on Evidence-Based Policymaking, set forth two key capacity challenges for the federal government related to data:

- The capacity to support the full range of evidence-building functions is uneven, and where capacity for evidence building does exist, it is often poorly coordinated within departments.
  - The federal evidence community has insufficient resources and limited flexibilities that restrict the ability to expand evidence-building activities.<sup>36</sup>
- A key recommendation of the Commission is that the President direct Federal departments to increase capacity for evidence building throughout government.

**Mark A. Abramson** is the President, Leadership Inc. and the founding Executive Director of the IBM Center for The Business of Government. His recently published books include *Getting It Done: A Guide for Government Executives* (with Daniel J. Chenok and John M. Kamensky) and *Succeeding as a Political Executive: 50 Insights from Experience* (with Paul R. Lawrence). He is also the author of *What Government Does: How Political Executives Manage* (with Paul R. Lawrence).

### Endnotes

- 1 Meredith Somers, "Federal managers working on 'culture change' for cross-component data sharing," *Federal News Radio* (April 26, 2017).
- 2 Somers, "Federal managers working on culture change."
- 3 For additional information on the President's Management Agenda, see <https://www.performance.gov/PMA/>.
- 4 Robert D. Behn, *The PerformanceStat Potential: A Leadership Strategy for Producing Results* (Washington, DC: Brookings Institution Press, 2014).
- 5 Paul O'Connell, *Using Performance Data for Accountability: The New York City Police Department's CompStat Model of Police Management*, IBM Center for The Business of Government, 2001.
- 6 Lenneal Henderson, *The Baltimore Citi-State Program: Performance and Accountability*, IBM Center for The Business of Government, 2003.
- 7 Christopher Patusky, Leigh Botwinik, and Mary Shelley *The Philadelphia SchoolStat Model*, IBM Center for The Business of Government, 2007.
- 8 Patusky, et al., *Philadelphia SchoolStat*, 32-33.

- 9 Harry P. Hatry, Elaine Morley, Shelli B. Rossman, and Joseph S. Wholey, *How Federal Programs Use Outcome Information: Opportunities for Federal Managers*, IBM Center for The Business of Government, 2004.
- 10 Shelley Metzenbaum, *Using State Information: Measuring and Improving Program Performance*, IBM Center for The Business of Government, 2003.
- 11 Tom Davenport and D.J. Patil, *Strategic Use of Analytics in Government*, IBM Center for The Business of Government, 2008, 6.
- 12 Davenport and Patil, *Strategic Use of Analytics*, 28.
- 13 Partnership for Public Service, *From Data to Decisions*, 2011, 3.
- 14 Partnership for Public Service, *From Data to Decisions II: Building an Analytics Culture*, 2012.
- 15 Partnership for Public Service, *From Data to Decisions III*, 2013.
- 16 Partnership for Public Service, *From Data to Decisions III*, 4.
- 17 Jennifer Bachner, *Predictive Policing: Preventing Crime with Data and Analytics*, IBM Center for The Business of Government, 2013, 14.
- 18 Sukumar Ganapati, *Using Geographic Information Systems to Increase Citizen Engagement*, IBM Center for The Business of Government, 2010.
- 19 Sukumar Ganapati, *Use of Dashboards in Government*, IBM Center for The Business of Government, 2011.
- 20 Ganapati, *Use of Dashboards*, 15.
- 21 Genie Stowers, *The Use of Data Visualization in Government*, IBM Center for The Business of Government, 2013.
- 22 Stowers, *Data Visualization*, 8.
- 23 Kevin Desouza, *Realizing the Promise of Big Data: Implementing Big Data Projects*, IBM Center for The Business of Government, 2014.
- 24 Desouza, *Realizing the Promise*, 10.
- 25 Alfred Tat-Kei Ho with Bo McCall, *Ten Actions to Implement Big Data Initiatives: A Study of 65 Cities*, IBM Center for The Business of Government, 2016, 6.
- 26 For more detailed discussions of the Recovery Act, see Francisca M. Rojas, *Recovery Act Transparency: Learning from State's Experience*, IBM Center for The Business of Government, 2012; and Earl Devaney, "Using Predictive Analysis to Prevent Rather than React and Respond: A Case Study of the Recovery Accountability and Transparency Board," *Fast Government: Accelerating Service Quality While Reducing Cost and Time*, IBM Center for The Business of Government, 2013.
- 27 The Pew Charitable Trusts, *How States Use Data to Inform Decisions*, 2018.
- 28 Pew Charitable Trusts, *How States Use Data*, 23.
- 29 Office of Management and Budget, *M-10-06: Open Government Directive*, December 8, 2009.
- 30 Office of Management and Budget, *M-13-13: Open Data Policy – Managing Information as an Asset*, May 9, 2013.
- 31 Office of Management and Budget, *M-14-06: Guidance for Providing and Using Administrative Data for Statistical Purposes*, February 4, 2014.
- 32 *Digital Accountability and Transparency Act of 2014*, Public Law 113-101 (May 9, 2014).
- 33 Jane Wiseman, *Data-Driven Government: The Role of Chief Data Officers*, IBM Center for The Business of Government, 2018
- 34 Francisca M. Rojas, *Recovery Act Transparency: Learning from States' Experience*, IBM Center for The Business of Government, 2012.
- 35 The Foundations for Evidence-based Policymaking Act, H.R. 4174, was introduced in October, 2017, 115th Congress (2017-2018).
- 36 Commission on Evidence-based Policymaking, *The Promise of Evidence-Based Policymaking*, 2017.