# The Impact of Big Data on Government

THE INTELLIGENT ECONOMY

Advances in technologies and the increasing amount of information are transforming how business is conducted in many industries, including government. Government data generation and digital archiving rates are on the rise due to the rapid growth of mobile devices and applications, smart sensors and devices, cloud computing solutions, and citizen-facing portals. As digital information expands and becomes more complex, information management, processing, storage, security, and disposition become more complex as well. New capture, search, discovery, and analysis tools are helping organizations gain insights from their unstructured data. The government market is at a tipping point, realizing that information is a strategic asset, and government needs to protect, leverage, and analyze both structured and unstructured information to better serve and meet mission requirements. As government leaders strive to evolve data-driven organizations to successfully accomplish mission, they are laying the groundwork to correlate dependencies across events, people, processes, and information. Figure 1 illustrates IDC's intelligent economy concept, with intelligent industry solutions deployed on the 3rd Platform where technologies drive the explosion of growth of digital information such as the information collected and disseminated from trillions of devices such as smartphones and embedded sensors.

The Intelligent Economy
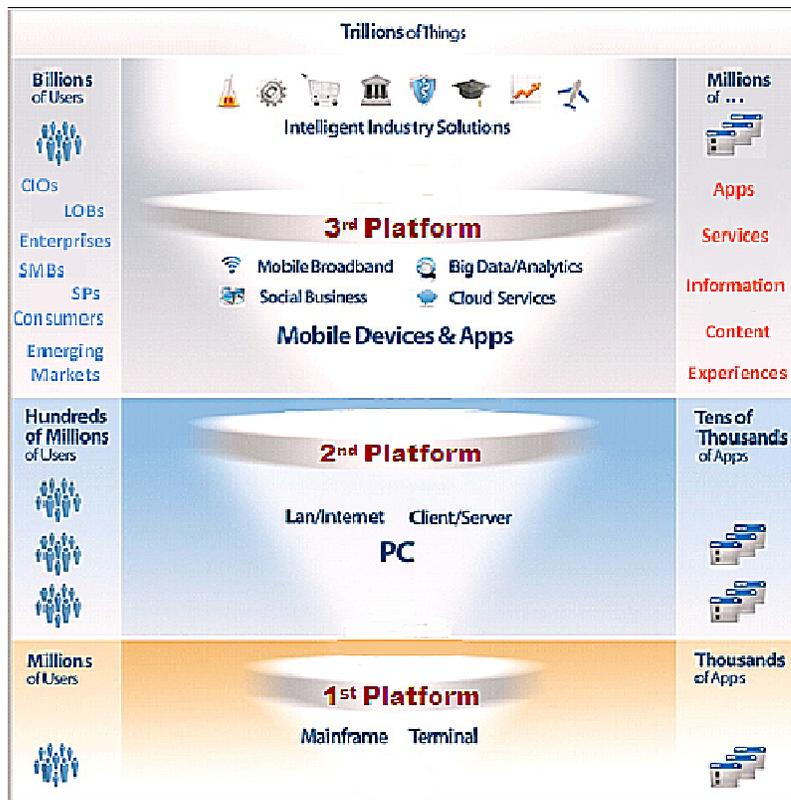
Source: IDC Government Insights, 2012

High-value government solutions will be created from a mashup of the most disruptive technologies:

● Mobile devices and applications

● Cloud services

● Social business technologies and networking

● Big Data and analytics

The use of laptops, smartphones, and tablets in government continues to grow. Mobile computing enables effective telework and supports continuity of operations and worker productivity in disaster recovery. The communication technologies of social media business technologies are enabling citizens to take a proactive role in government. As social media skills are increasingly well honed, the value of this collaborative technology is maturing, especially when used to enable open, transparent government and to facilitate government service delivery.

IDC predicts that by 2020, the IT industry will reach $5 trillion, approximately $1.7 trillion larger than today, and that 80% of the industry's growth will be driven by these 3rd Platform technologies.

In the long term, these technologies will be key tools for dealing with the complexity of increased digital information. Big Data is one of the intelligent industry solutions and allows government to make better decisions by taking action based on patterns revealed by analyzing large volumes of data — related and unrelated, structured and unstructured.
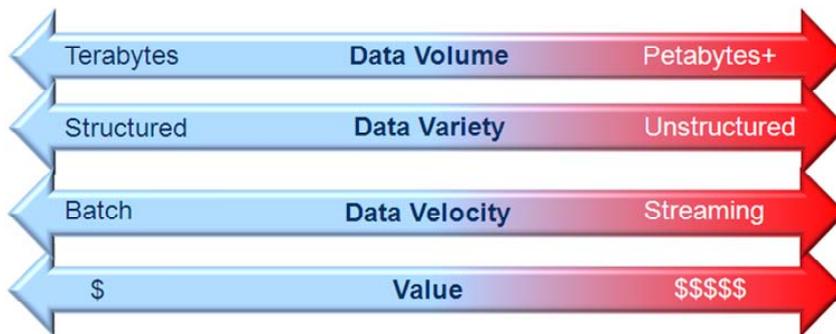
BIG DATA

IDC's research shows that Big Data technologies are at the heart of the intelligent economy and the solutions that enable it. IDC defines Big Data technologies as *a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high-velocity capture, discovery, and/or analysis.* New capture, search, discovery, and analysis tools can help organizations gain insights from their unstructured data, which accounts for more than 90% of the digital universe.

IDC's Big Data definition encompasses hardware, services, and software that integrate, organize, manage, analyze, and present data that is characterized by volume, variety, velocity, and value extracted and leveraged from the data to further mission (see Figure 2).

FIGURE 2

IDC's Big Data Technology Definition

Big Data technologies describe a **new** generation of technologies and architectures designed to **economically** extract **value** from very large **volumes** of a wide **variety** of data by enabling high-**velocity** capture, discovery, and/or analysis.

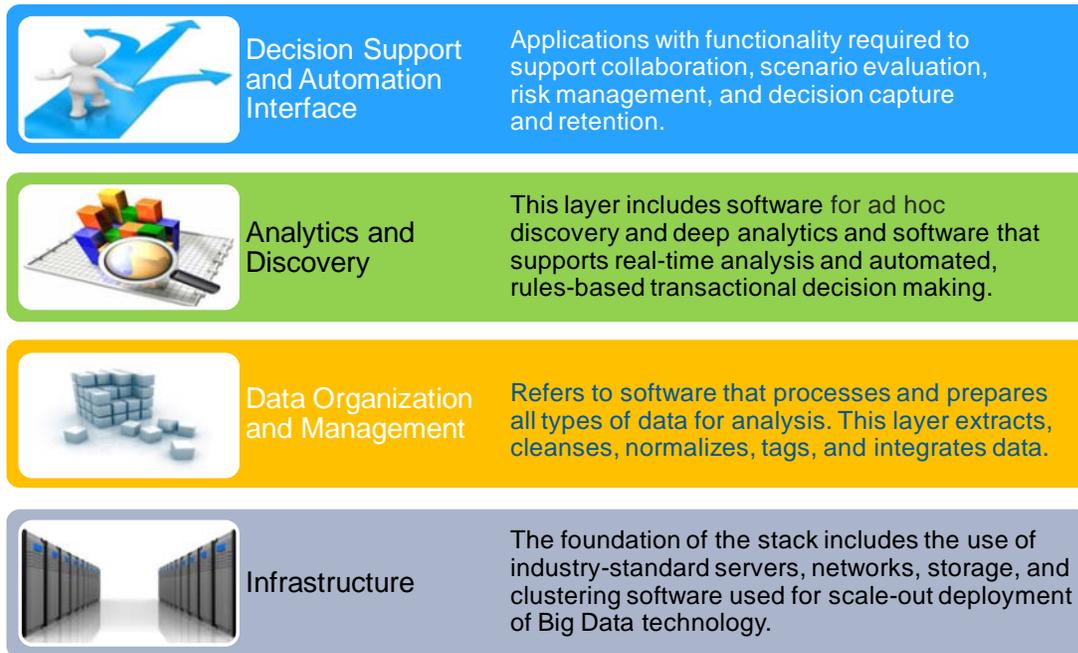| Terabytes | Data Volume | Petabytes+ |
Structured	Data Variety	Unstructured
Batch	Data Velocity	Streaming
$	Value	$$$$$

Source: IDC, 2012

Figure 3 shows IDC's Big Data technology stack. The foundation of the stack includes the use of industry-standard servers, networks, clustering software used for scale-out deployments of Big Data technology, and storage. Storage is becoming a strategic asset as the explosion of structured and unstructured information drives concerns about backup, recovery, and archiving in government. Tight budgets coupled with ever-growing government information and a move to digital information make it critical for government agencies to rightsize their traditional and Big Data information storage. Many agencies are rethinking the roles of traditional storage and deploying a cost-efficient and combined approach of tape for archival storage, disks for often requested records, and cloud storage for Big Data information.

FIGURE 3

IDC's Big Data Technology Stack

	Decision Support and Automation Interface	Applications with functionality required to support collaboration, scenario evaluation, risk management, and decision capture and retention.
	Analytics and Discovery	This layer includes software for ad hoc discovery and deep analytics and software that supports real-time analysis and automated, rules-based transactional decision making.
	Data Organization and Management	Refers to software that processes and prepares all types of data for analysis. This layer extracts, cleanses, normalizes, tags, and integrates data.
	Infrastructure	The foundation of the stack includes the use of industry-standard servers, networks, storage, and clustering software used for scale-out deployment of Big Data technology.

Source: IDC, 2012

GOVERNMENT DRIVERS INFLUENCING MARKET TRENDS

Easy and timely retrieval and analysis of related and unrelated information is crucial for government to meet and improve mission requirements that are varied across agencies. Data continues to be generated and digitally archived at increasing rates driven by Open Government initiatives, sensors, citizen interactions, and program transactions. Government organizations are beginning to deploy

Big Data technologies to analyze massive data sets in science and research as well as mine data to prevent bad actors from committing acts of terror and/or to prevent waste, fraud, and abuse. Drivers impacting Big Data in government are discussed in the following section.

Big Data Research and Development Initiative

In March 2012, the federal government launched the Big Data Research and Development Initiative with $200 million in new spending to improve the tools/techniques needed to track, access, organize, store, model, and analyze information and glean discoveries from huge volumes of digital data. This initiative is focused on government's use of Big Data for scientific discovery, environmental and biomedical research, education, and national security and includes the following:

- The Department of Defense (DOD) is investing approximately $60 million annually for new projects that will harness and utilize massive data in new ways and bring together sensing, perception, and decision support to make truly autonomous systems that can learn from experience, maneuver and make decisions on their own, and understand the limits of their knowledge. DOD is also planning to improve situational awareness to help warfighters and analysts and provide increased support to operations.

- The Defense Advanced Research Projects Agency (DARPA) is beginning the XDATA program, with $25 million annually for four years to develop computational techniques and software tools for analyzing large volumes of data, both semistructured and unstructured (text documents and message traffic).

- The National Institutes of Health (NIH) and the National Science Foundation (NSF) are investing in Big Data science and engineering. This research is focused on managing, analyzing, visualizing, and extracting useful information from large data sets; the NIH is particularly interested in those relating to health and disease — molecular, cellular, electrophysiological, chemical, behavioral, epidemiological, and clinical.

- The Department of Energy will provide $25 million in funding to establish the Scalable Data Management, Analysis, and Visualization (SDAV) Institute. Led by Lawrence Berkeley National Laboratory, the SDAV Institute will bring together the expertise of six national laboratories and seven universities to develop new tools to help scientists manage and visualize data on the department's supercomputers.

- The U.S. Geological Survey (USGS) launched Big Data for Earth System Science. This initiative will improve understanding of species response to climate change, earthquake recurrence rates, and the next generation of ecological indicators.

Building a 21st Century Digital Government

In May 2012, President Obama issued a directive entitled *Digital Government: Building a 21st Century Platform to Better Serve the American People* with a 12-month action plan aimed at providing better services to citizens by delivering anytime, anywhere digital information and services to federal workers on any device, safely and securely. The strategy requires agencies to use machine-readable information, mobile information, and shared services. This strategy provides key principles for guiding federal IT decision makers, including:

- **Information centric.** Agencies are to make open data, content, and Web application programming interfaces (APIs) the new default and make existing high-value data and content available through Web APIs to create a more information- and customer-centric government. For example, the U.S. Census Bureau recently released an API that opens access to census data and gives citizens better access to important U.S. statistics such as demographic, socioeconomic, and housing statistics.

- **Shared platform.** The federal government has established a Digital Services Innovation Center and Advisory Group to provide intra-agency support to "build once, use many times" and shift to an enterprisewide asset management and procurement model and accelerate innovative digital services.

- **Customer centric.** Agencies are to measure performance and customer satisfaction to improve service delivery, apply metadata tagging to two major customer-facing systems by June 2013, and deploy modern tools and technologies to deliver better digital services.

Open Government Plans

In their revised Open Government 2.0 plans, agencies were required to address their plans for using Big Data. For example, the U.S. Department of Health & Human Services, with an estimated waste, fraud, and abuse cost of approximately $66 billion in 2011 (see paymentaccuracy.gov), created plans that look into new insights and interactions of linked data sets to better understand healthcare expenditure, use of services, and cost of care at the community and provider levels.

Presidential Memorandum

The Presidential Memorandum for Managing Government Records instructs agencies to transition to electronic records as a way to save taxpayer dollars, promote accountability, and increase transparency. Additionally, the August 24, 2012, OMB Memorandum requires agencies to designate a senior agency official (SAO) with broader agencywide responsibilities to oversee agency records management. The SAO must be a senior official at the Assistant Secretary level or its equivalent with responsibility for ensuring that the agency efficiently

and appropriately complies with all applicable records management statutes, regulations, and policies and ensures that the following timelines are met:

- **By December 31, 2016:** Federal agencies must manage all email records, both permanent and temporary, in an accessible electronic format. Email records must be retained in an appropriate electronic system that supports records management and litigation requirements, including the capability to identify, retrieve, and retain the records for as long as they are needed.

- **By December 31, 2019:** All permanent electronic records in federal agencies will be managed electronically to the fullest extent possible for eventual transfer and accessioning by the National Archives and Records Administration (NARA) in an electronic format. Agencies should consider the benefits of digitizing permanent records created in hardcopy format or other analog formats (e.g., microfiche, microfilm, analog video, and analog audio).

Many Decision Makers with Different Drivers

Big Data represents a big opportunity and big challenges for government executives given that today's transaction-based systems and siloed programs discourage information sharing and hinder movement from process driven to data driven. Government leaders often have different drivers for using Big Data. For example, program directors and managers want access to Big Data to improve program accuracy, better serve constituents, and protect the warfighter. As chief financial officers shift their focus from asset management to agency outcomes, they want to establish key linkages of financial and nonfinancial information and use business intelligence for decision support. Inspectors general want to deploy preventive measures to prevent billions of dollars from being spent on waste, fraud, and abuse. Chief information security officers want to prevent cyberattacks, protect critical information, and overcome obstacles in identity management. Chief information officers (CIOs) need to support FOIA, records management, and eDiscovery, all while driving down costs, adhering to policy and regulations, and supporting the use of Big Data by other agency executives and directorates.

As government leaders strive to create data-driven organizations to successfully accomplish mission, CIOs are laying the groundwork to correlate and track dependencies across people, processes, and information to ensure that the right information is available to the right person at the right time. CIOs and records managers need to work with SAOs and deploy technology and services to better manage large volumes of structured and unstructured data to provide insight into operations and mission performance and provide relevant, timely information for better decisions. Agencies need to take a strategic approach to where they are placing all this data and on the medium that provides the greatest success of access when needed.

Challenges to Adopting Big Data and Managing the Life Cycle of Big Data in Government

Government faces many challenges in managing the life cycle of Big Data as government's traditional silo approach hinders sharing knowledge and working across organizational boundaries. Government also processes multiple types of information (paper, digital, multimedia). IDC Government Insights research shows that if paper documents were available in digital format, government workers would gain almost an hour a day. In addition, nearly half of government digital documents are printed, and almost 40% of these printed documents are reentered as digital documents. But simply digitizing information isn't enough. IDC Government Insights research also shows that a fourth of government employees were unable to find or access the digital information they needed more than half the time, indicating that making information available in digital formats is not enough and that records need to be indexed and properly tracked in order to be accessed.

IDC Government Insights recently attended a summit of approximately 30 government CIOs. When we asked the executives about their single biggest challenge in dealing with increased data, they identified analytics and storage as their top challenges, as shown in Figure 4. Government executives recognize the importance of gaining insights through the proper storage, retrieval, and analysis of Big Data. The need of government decision makers to develop, deploy, and manage an effective Big Data environment is driving many to look to information storage providers to support their current and future needs.

FIGURE 4

Government Challenges in Dealing with Increased Data
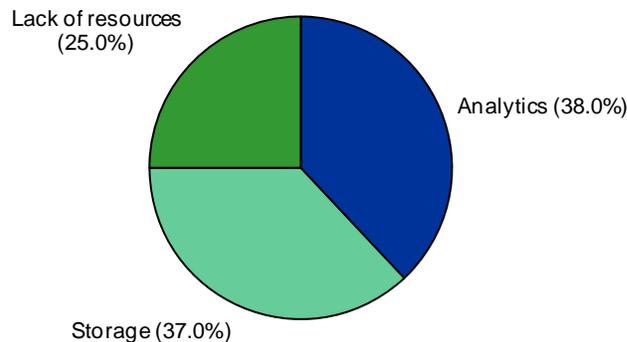
Source: GDS Government CIO Conference Participants, June 2012

#GI237315 ©2012 IDC Government Insights

A survey that IDC and *Computerworld* conducted across multiple industries indicates that many resource and skills shortages challenge the adoption of Big Data. As shown in Figure 5, 68.9% of respondents are looking for advanced analytics/predictive analytics, the number 1 skill challenge. However, survey respondents require skills across many areas, including industry or business process expertise, rules management, and complex event processing. The availability and the skill level of Big Data information management talent will have a direct impact on government's ability to successfully deploy Big Data.

FIGURE 5

Resource/Skills Shortage a Big Data Challenge

Q   What type of skills are you looking for most over the next 12 months?

(% of respondents)
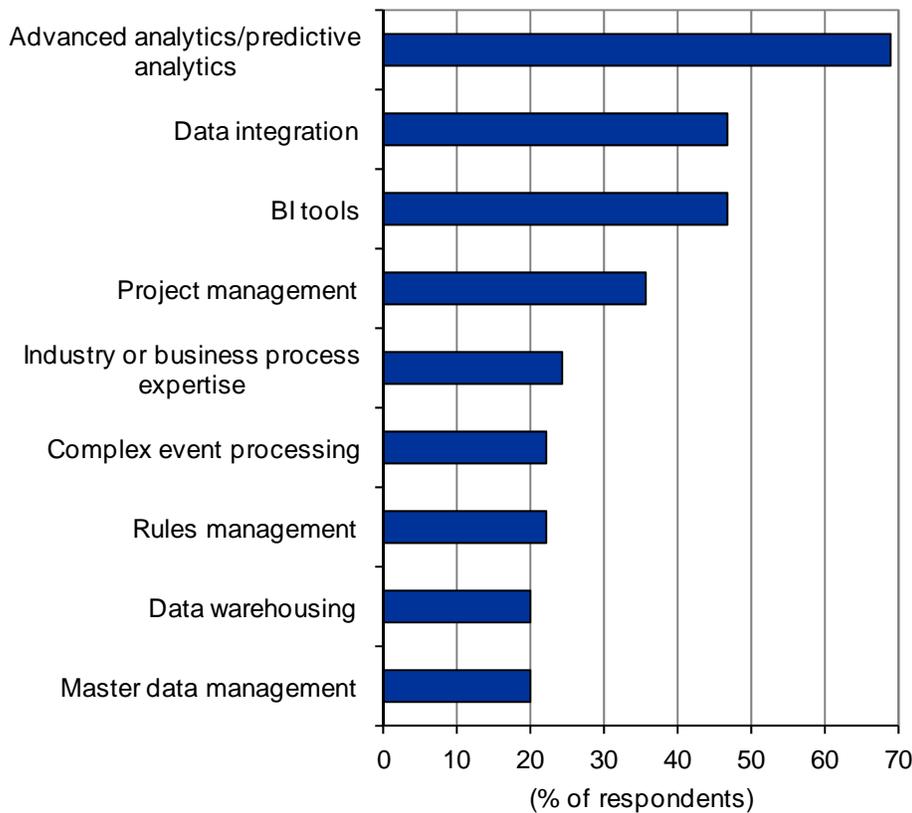
n = 45 services firms

Source: IDC and *Computerworld* Business Intelligence and Analytics Survey, 2012

RECOMMENDATIONS TO GOVERNMENT

Government agencies should explore economically viable automated records management technologies and solutions to reduce the burden of records management responsibilities. Technology is continuously evolving to provide advanced solutions to support data backup, recovery, and archiving requirements. Storage infrastructure that is capable of addressing the specific demands around data volume, velocity, and variety (three of the four "Vs," as described in IDC's Big Data framework) will be critical for government agencies to successfully use and process Big Data. Greater visibility into organizational information — together with the ability to produce documentation when needed — allows agencies to enhance employee efficiency and productivity, respond to questions, make informed decisions by connecting the related and unrelated information, and provide the information needed for transparency, collaboration, and participatory government.

Agencies should also strategically respond to the growth of data with plans that address the storage of traditional data as well as the storage of digital information — structured and unstructured and metadata. Budgetary constraints are forcing agencies to rightsize their Big Data information storage. Plans should include the growing roles of tiered storage, tape, and cloud-based backup and recovery as well as an analysis of the cost versus the benefit of multiple storage choices.

Tape provides efficient, cost-effective, and reliable storage for nonactive, rarely requested data. It allows government agencies to keep valuable enterprise data secure on tape media. The majority of government information is likely inactive. Tape offers long-life, high-capacity storage for long-term data retention to support compliance as well as archiving of nonactive data.

Tape-based backup and recovery offers a more reliable and economical option than disk-based solutions, but it requires more hands-on management. Tiering data storage with disk and tape versus only disk storage can provide a more cost-effective storage medium for nonactive data, enabling agencies to optimize onsite and offsite backup processes.

Cloud-based backup, recovery, and archiving provide different benefits and can play a complementary role in a storage strategy. For data that is constantly changing and requires short real-time access, cloud technology provides an excellent backup and recovery solution. It enables movement of data offsite faster and more frequently and virtually instant retrieval. While cloud-based solutions can be more expensive than tape over the long term, they offer fast, automated backup and archiving that helps reduce IT's workload. Using cloud technology for backup, recovery, and archiving activities offers a flexible, scalable, pay-as-you go approach to help keep costs down while minimizing resources required to manage related processes.

Many agencies are rethinking the role of traditional storage and deploying a cost-efficient and combined approach of tape for archival storage, disks for often requested records, and cloud storage for Big Data information. A strategic approach to storage allows agencies to produce documentation when needed, enhances employee efficiency and productivity, and provides the information needed for transparency, collaboration, and participatory government.

IRON MOUNTAIN STRENGTHS AND CHALLENGES

While emails, instant messages, data files, document files, and scanned images are all driving the growth of Big Data, managing and storing this information — and its growth — are not trivial tasks. Processes that were once simple with paper records — such as indexing, retrieval, retention management, and destruction — are now complex. When government entities look to modernize agency records management programs and respond quickly to accessing, using, and analyzing information, it's crucial that they partner with a trusted provider with a history of expertise in records and information management.

Iron Mountain offers a product and services portfolio that is consistent with what IDC Government Insights expects from a market-leading technology provider to the government segment. Iron Mountain has decades of experience in providing records management and physical storage for both physical and digital assets. The company's subject matter experts take a consultative approach to individual agency environments to determine the best opportunities for an efficient and cost-effective way to manage and store government records, including:

● Digitizing paper records to an electronic format to improve information accessibility, sharing, and collaboration

● Digital archiving, email archiving, and image archiving services including solutions for email management and inactive file storage that improves control and management abilities while avoiding tying up capital

● Combining paper files, digital information, and hybrid files in a manner consistent with each agency's unique needs

● Gaining access to tools to organize data for easy access for information requests and helping agencies comply with the Open Government Directive, the Digital Government Strategy, and the Presidential Memorandum for Managing Government Records

● Providing cost-efficient storage options that combine, tape, disk, and cloud storage to allow government agencies to rightsize their traditional and Big Data information assets

Iron Mountain faces several unique market challenges and many of the same market challenges as other information management vendors servicing the government Big Data marketplace, including:

- **Demonstrating leadership in the transition from information management to Big Data.** While Iron Mountain's legacy records management and storage solutions are not broadly implemented across the government sector, they have made inroads. As Iron Mountain continues this momentum, the company must more effectively demonstrate leading capabilities at both the records management level and the storage level and provide high-profile wins.

- **Offering competitive, differentiated solutions portfolios.** The government market is beginning to explore the value of Big Data and is budgeting accordingly. Iron Mountain's competitors are also focused on expanding their solution portfolios on the Big Data trend in terms of breadth and depth of product capabilities, storage solutions, and professional services. Iron Mountain must leverage its nongovernment industry experience, offer a variety of best practice suggestions derived from lessons learned thus far, and demonstrate the value that it can bring to government to gain Big Data market share in this industry.

- **Focusing on security.** Because of the sensitive nature of government data and the requirements and penalties imposed by government regulations, security is top of mind for government IT executives. Service providers, such as Iron Mountain, that supply their private sector customers with physical and digital information protection, access, retention, storage, and disposal need to emphasize not only their leadership in the private sector providing comprehensive records management solutions but also their end-to-end information security plans and execution road maps that serve the unique demands of the government sector.