THE GOVLOOP GUIDE

TRANSFORMING YOUR AGENCY WITH BIG DATA

WELCOME / CONTENTS

FOREWARD

Government technology decision makers in all three branches of federal government and from many States consistently tell me about an issue they have with "Big Data". This is an issue you have seen too I'm sure. The issue is a lack of discipline and rigor over the use of the term itself. What began as a concept describing a specific set of new capabilities to enable analysis that could never be done before has become a movement where every strategist, marketer and pundit seeks to offer their own definition and use of the term. That is why I was so happy to see GovLoop's mission-focused approach to this issue. By focusing on the missions of enterprises and then surveying real decision-makers for issues they are helping bring discipline and focus to the domain.

Why is discipline important at this stage? Do you remember service-oriented architecture (SOA)? This concept led to tremendous new capabilities and efficient, mission-focused designs. Enterprises established architectures in which application interfaces, logic and data were separated and smartly reusable. After the term went mainstream, every company in the IT ecosystem grabbed onto it and began to use the acronym SOA to mean anything they wanted it to. Although it's still a useful construct for IT professionals, when it comes to interacting with industry, the term has now lost much of its meaning.

If we let the term "Big Data" go the way of "SOA" we will sub-optimize its use. If we keep a mission-focused discipline on the term we will help drive continuing improvement and optimize the insertion of new innovative technologies into governments.

As you dive into this report keep the following definition of Big Data in mind. It flows from Doug Laney's famous articulation of the "Three V's" and is consistent with the collaborative efforts of the TechAmerica foundation and with our reporting at CTOvision.com and with the community edited WikiPedia entry for Big Data:

> **Big Data:** A phenomenon defined by the rapid acceleration in the expanding volume of high velocity, complex and diverse types of data. Big Data is often defined along three dimensions– volume, velocity and variety.

> **Big Data Solutions:** Advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of information. A Big Data solution uses new approaches for sense-making over large quantities of data, such as in the Apache Hadoop Big Data Platform.

Please keep these terms in mind as you review this Govloop guide to Transforming Your Agency with Big Data. The insights you find here will help you learn from others who are also on the journey and will also help you speed along the Big Data projects you may already have underway.

BOB GOURLEY
http://ctovision.com
http://crucialpointllc.com
bob@crucialpointllc.com
703 994 0549

EXECUTIVE SUMMARY

The way agencies store, manage and collect data is changing. In the past, agencies analyzed data from stable enterprise services. Data was created from customer relationship management services, enterprise resource planning systems or from the variety of financial instruments agencies used to process transactions. Agencies could then run common reports, store data in warehouses or databases, and easily extract knowledge to inform their decision-making. With the explosion of data from social media, the Internet and growing examples of online transactions, government has more opportunities than ever before to transform agencies through big data analysis.

Through qualitative and quantitative analysis, this report will highlight the common challenges, best practices and pertinent issues surrounding big data. ***Transforming Your Agency Through Big Data*** includes a survey from 215 public sector professionals, and interviews with government and industry thought leaders. Expert interviews include:

NIKKI CLOWERS, *Director of Financial Markets and Community Investment Issues at GAO*

MARK HEADD, *Chief Data Officer, City of Philadelphia*

AUDIE HITTLE, *Chief Technology Officer (CTO), Federal Market at EMC Isilon*

JENNIFER KERBER, *President of the TechAmerica Foundation*

SHAWN KINGSBERRY, *Chief Information Officer, Recovery Accountability and Transparency Board (RATB)*

PAT HERBERT, *Principal Solutions Architect, SAS*

Our report identifies that government agencies are still at the early stages of adopting big data initiatives. To fully leverage big data analysis, agencies are in need of guidance and best practices to fully capitalize on the data that they are collecting. In our survey, GovLoop found that education (19%), acquiring infrastructure (18%), and confidentiality and access control (16%) were some of the challenges facing government.

Additional findings highlight the ways government has leveraged big data. The survey finds that budget/finance (14%), communications/public affairs (14%) and health and human services (11%) were the leading mission areas in which organizations have leveraged big data analysis. Additionally, the survey finds that the majority of organizations have yet to hire a data scientist, as 61% of respondents do not have data scientist on staff. To begin to craft big data strategies, GovLoop believes that agencies should:

1. Consider parallels and lessons learned from existing open data initiatives.

2. Identify resource needs and staffing.

3. Be strategic and develop clear vision statement and goals.

4. Work diligently to extract knowledge from legacy systems and unstructured data.

5. Government to government collaboration is essential, agencies need to focus on standardizing data and developing shared infrastructures.

This report concludes with a big data cheat sheet, as a means to quickly get the 'need to know' facts on big data and a brief outline of the key findings in this report. As budgets continue to plague the public sector, agencies are searching for new efficiencies and innovative ways to find actionable insights from big data analysis. This report can serve as the first step for your agency to harness the power of data, and improve decision-making, information sharing and service delivery of the agency through data analysis.

THE STATE OF BIG DATA

Highlights from the GovLoop Survey

Recently, GovLoop conducted a survey on big data applications across government. The survey included 215 responses primarily from a federal audience (41%), and also included participants from state level employees (24%), local government (23%) and contributions from academia and industry members (12%). Throughout this report, GovLoop shares additional findings from our survey.

One question explored, "Do you or your organization use big data in a way that has an impact on any of the following mission areas?" Respondents could choose from acquisition/contracting, budgeting/finance, communications/public affairs, economic development, emergency management, environmental protection, health and human services, parks and recreation, public works, property management and transportion. The results can be found in Figure 1. Respondents indicated the top three results for big data applications were for budgeting/finance (14%), communications/public affairs (14%), and health and human services (11%). Respondents also had the chance to submit other ways they are leveraging big data analysis on mission areas, participants identi-

fied areas such as performance measurement, security and defense and education and skill development.

Kimberly Denz, Business Systems Analyst, Mississippi Department of Marine Resources, provided some ways that big data is starting to be leveraged by the State of Mississippi for mission centric activities. Denz provided the example of monitoring coastal ecology and tracking compliance. Denz states that it is difficult to know if the agency is out of compliance because of how many permits have been issued, which can be capped to preserve coastal plains and the environment. Another scenario Denz identifies is leveraging existing data across agencies. The scenario she provides is imagining if the Mississippi Marine Patrol cites an individual for a violation, they can then check and know if the individual is wanted by the FBI, and proceed appropriately.

Additional questions explored where agency data came from (Figure 2), and who creates agency data (Figure 3). The survey shows that the leading way respondents collect data is through capturing information from documents, whether they are print or digital. Clearly, with the growth of self service platforms, and increasing examples of transactions performed online, government is flooded with documents, and they serve as one of the primary sources of data for organizations. Additionally, web data and email both were in the top three for data sources.

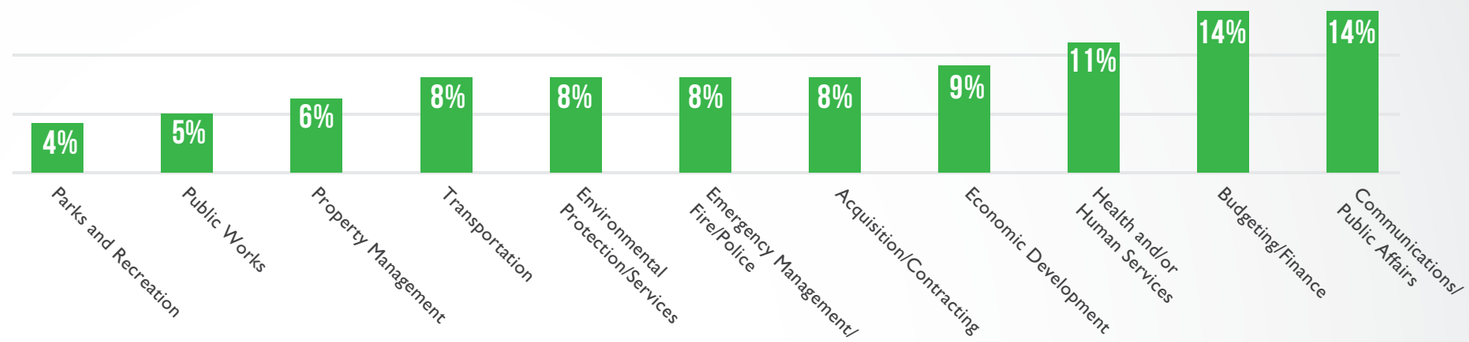
As more services move to the web, these findings are consistent with the emerging ways constituents engage with government. Respondents stated that data is created through direct collection strategies, environmental monitoring and research, licensing and geospatial information systems (GIS).

Figure 3, which highlights data creators, shows that most of the data created by government is from internal stakeholders, as in employees, from internal processes or through various inter-governmental collaborations. Some of the key results from figures 3 and 4 are:

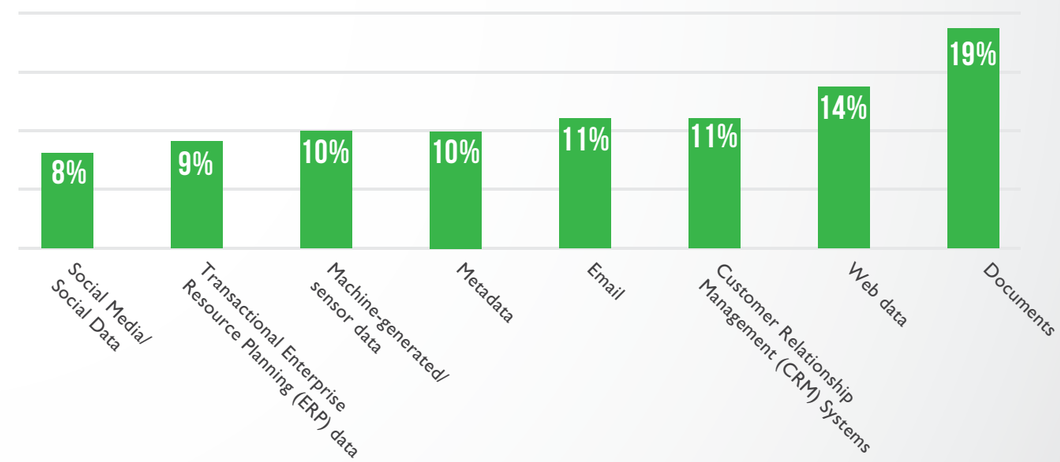
- Data that government collects is varied
- Need to find insights from multiple sources of data
- Standardization continues to be a challenge
- Legacy systems make it challenging to access historical data

BIG DATA IMPACT ON MISSION AREAS (FIGURE 1)

DO YOU OR YOUR ORGANIZATION USE BIG DATA IN A WAY THAT HAS AN IMPACT ON ANY OF THE FOLLOWING MISSION AREAS? (PLEASE SELECT ALL THAT APPLY)



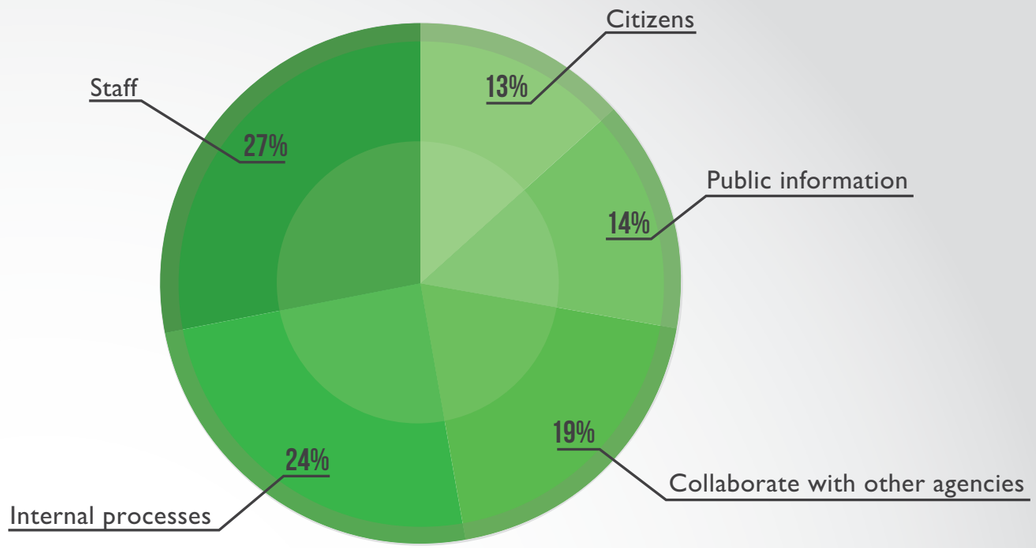
WHERE DOES YOUR AGENCY DATA COME FROM? (FIGURE 2)



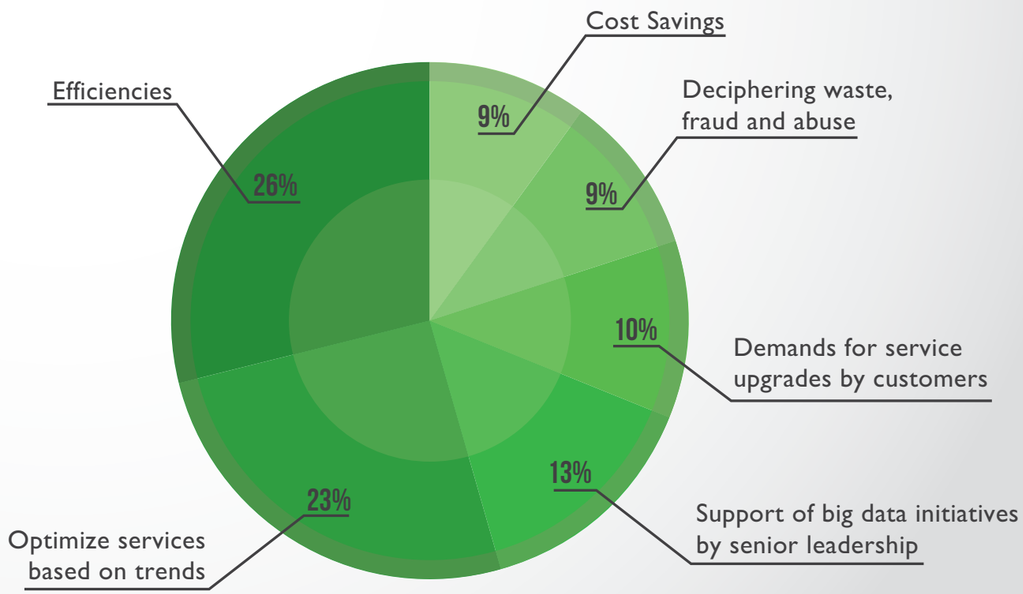
The survey also explores what the core drivers for big data adoption has been. Our survey finds that the main reason government is adopting big data initiatives is due to increasing needs to find efficiencies in government. Following efficiencies is the desire to optimize services based on trends. Clearly, government organizations are looking to use big data to improve the service delivery for citizens, and identify new and emerging ways to optimize and streamline processes. Survey respondents mentioned, they have adopted big data programs for "holistic perspective of complex issues," and "Support of increased analytical capabilities by organization."

The GovLoop survey provided a snapshot of how agencies are planning to use, or how they are using big data analysis. In many cases, this is uncharted territory for government agencies. In the following sections of this report, GovLoop spoke with government and industry leaders who are leading the way with big data analysis. Through their expert insights, government agencies can learn the best practices and craft strategies to help transform their agency through big data.

WHO CREATES YOUR AGENCY'S DATA? (FIGURE 3)



WHAT HAS BEEN THE CORE DRIVER FOR BIG DATA INITIATIVES? (FIGURE 4)



LEADING THE WAY

Big Data Lessons from Philadelphia

EXPERT INSIGHTS WITH MARK HEADD, CHIEF DATA OFFICER, CITY OF PHILADELPHIA

In August 2012, Philadelphia Mayor Michael Nutter named Mark Headd as the City's first Chief Data Officer. In this role, Headd is responsible for overseeing open data and transparency initiatives across Philadelphia. In this report, Headd provides his expert insights on how municipal governments can leverage big data, the relationship between big data and open data, and identifies lessons learned and insights for agencies to craft big data strategies.

The City of Philadelphia is still exploring how to effectively leverage big data to improve citywide operations. Headd notes that the city has started several pilot programs that combine city data with data obtained from various social networks and internal resources. Headd states, "Along with most cities, we are very much at the very beginning stages of adopting big data programs." In many cases, big data programs are still in their nascent stages, and this section of the report highlights Headd's expert insights on how to leverage data to improve public sector decision-making.

Headd defines big data as, "the extraction of big ideas from data, regardless of the size of data." For the City of Philadelphia, big data is not as much about specialized technology, such as Hadoop and MapReduce, as it is about collaboration and extracting meaningful information from data across agencies and departments. "Big data is less about the size of

big data and more about the extraction of actionable intelligence of data," stated Headd.

Headd continues to describe the difference between federal and local applications of big data: "Philadelphia does not have petabytes of data, we are not the census bureau, we are not the federal government, at a municipal government scale, and even at the state level, it's less about specialized technology, and its more about data across departments and across agencies." Although Headd identifies that Philadelphia is not collecting the same volume of data as the federal government, across all levels of government, agencies are challenged to break down siloes and work collaboratively to share data, which can provide leaders with new insights and improved decision-making.

CHALLENGES TO ADOPTING BIG DATA AT THE CITY OF PHILADELPHIA

As the City of Philadelphia began to explore big data initiatives, the city faced two core challenges, extracting knowledge from legacy systems and a fragmented style of governance.

"THE PROBLEM IS THAT THE TOOLS FOR USING DATA HAVE DEVELOPED SO QUICKLY AND BECOME SO POWERFUL THAT THEY REALLY HIGHLIGHT THE FACT THAT WHEN IT INVESTMENT DECISIONS WERE MADE, THEY OFTEN WERE MADE WITHOUT THE FORESIGHT THAT SOMEDAY 'WE'D WANT TO DO DIFFERENT THINGS WITH THIS DATA,' AND I DON'T THINK THAT IS UNUSUAL ACROSS GOVERNMENT"

- Mark Headd, Chief Data Officer, City of Philadelphia

CHALLENGE 1: EXTRACTING KNOWLEDGE FROM OUTDATED LEGACY SYSTEMS

A legacy system is simply an antiquated technology or application that government uses. In many cases, outdated technology makes it very challenging for organizations to extract data and identify value. As Headd states, "people didn't even know what would be possible with modern systems. We simply didn't know what we didn't know." Headd identified that many of the technologies available today were not on anyone's radar when investment decisions were made. The challenge with legacy systems is that they make it challenging on a enterprise wide scale to:

- Extract data
- Share data
- Combine data
- Analyze date

"The problem is that the tools for using data have developed so quickly and become so powerful that they really highlight the fact that when IT investment decisions were made, they often were made without the foresight that someday 'we'd want to do different things with this data,' and I don't think that is unusual across government," states Headd.

CHALLENGE 2: SHARING DATA WITHIN A FRAGMENTED GOVERNMENT STRUCTURE

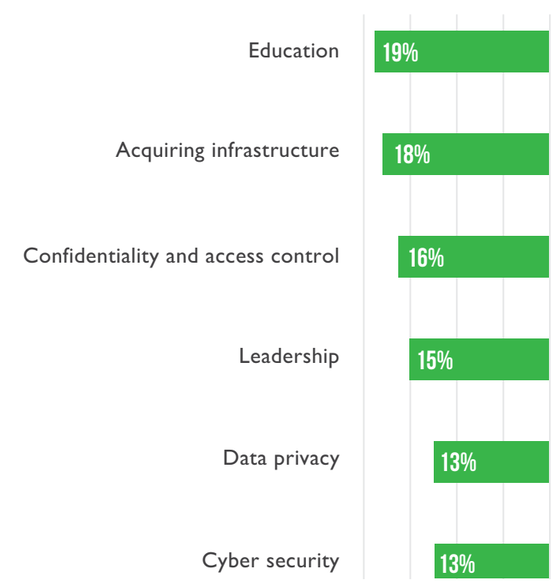
Typically, a fragmented government is used to describe the relationship between county and city government – with no central authority. Here we are using the term 'fragmented' to describe the multiple agencies and actors involved with data collection and service delivery within the City of Philadelphia. Within the City there are multiple government entities serving the needs of citizens. For instance, to name a few:

- Transportation authority that runs trains, buses and the subway
- School district that operates independently from the City of Philadelphia
- Separate parking authority that operates street and lot parking throughout the city
- Public safety programs

This means that there is substantial overlap of services by government agencies, creating multiple data sets. "Data often within the same government is often dispersed, and data is in different formats, in different systems and siloed," states Headd. For Headd, a core challenge is to combine data sets, bring in additional data, such as social media, and then work to extract actionable intelligence from data.

FROM THE GOVLOOP SURVEY: CHALLENGES TO IMPLEMENTING BIG DATA INITIATIVES (FIGURE 5)

Our survey explored challenges to big data. The data in Figure 5 shows that agencies are faced with a myriad of challenges on adopting big data, and still in the early stages of big data adoption, learning and working through challenges to many challenges to implement big data programs.



This is an enormous task for the City of Philadelphia, but is a core requirement for the city to advance big data programs. Headd is working to create a "shareable infrastructure," where agencies can easily access and leverage data citywide. This will be essential to capitalizing on the various kinds of data that the city collects, identifying new insights and improving the operations of government.

BEST PRACTICES TO IMPLEMENTING BIG DATA INITIATIVES

Although big data is just beginning in the City of Philadelphia, Headd provides his thoughts on what agencies can do to leverage big data programs. Like with most IT investments, one of the first steps is to clearly state the business problem, and how a robust use of big data will help achieve the organization's end goals. One of the interesting insights from Headd was the relationship between open data and big data, and from Headd's perspective, how open data initiatives can lead organizations to develop big data programs. Additionally, Headd asserts the importance of scaling services and offerings to meet future demand.

CONSIDER LESSONS FROM OPEN DATA INITIATIVES

During the interview, Headd explored how big data differs from open data. Headd's insights show that both are interconnected, but serve a different mission need for government agencies. While open data is used for outside consumers to make informed data based decisions, big data is used for organizations to make decisions at the enterprise level. For instance, an open data initiative may allow citizens to report potholes for the city transportation to fix, while a big data initiative may identify traffic patterns to learn what areas of the community need to be allocated funding to upgrade roads and infrastructure. Open data focuses on the government-to-citizen relationships, while big data focuses on government-to-government relationships.

Headd provides more clarity on the distinction between big data and open data: "Big data is more about organizations using data to inform their own decisions, where as open data is more about giving the public, outside consumers data to inform what they do, decisions they make."

Although big data and open data are slightly different, Headd recommends prior to starting a big data initiative, agencies should explore open data programs. "I always describe open data as the gateway drug to big data, because it makes government think about there data differently, open data makes government think more strategically about their data, it makes them think about the share-ability of data, which is one of the big hindrances to using big data and to using data to inform better decision making on the operational side of government," states Headd.

For the City of Philadelphia, big data is not as much about specialized technology, such as Hadoop and MapReduce, as it is about collaboration and extracting meaningful information from data across agencies and departments.

Open data and big data have a lot of parallels. In both cases, agencies are forced to think about how to leverage data to serve the most critical functions. Along the way, agencies must consider security implications, work with customers to meet needs (could be a citizen or an agency), and clearly identify success metrics.

3 STEPS TO LEVERAGE OPEN DATA TO IMPROVE BIG DATA PROGRAMS

1. Open data and big data have parallels to adoption

2. Process to open data to external clients is similar to leveraging data internally

3. Focus on the process and think strategically on data usage

"When agencies go through the process of opening data up for public consumption agencies have to go through a lot of the same steps that they would need to go through to share data internally, so it really sort of changes the way people think about data," states Mark Headd, Chief Data Officer, City of Philadelphia

3 INFRASTRUCTURE NEEDS FOR BIG DATA

1. Scalable offerings to meet increasing demands from stakeholders

2. Shareable infrastructure to facilitate collaboration across departments

3. Re-engineer systems to upgrade legacy systems to access historical data

BUILD INFRASTRUCTURES THAT SCALE AND ARE EASILY SHARED

Headd noted that agencies did not prepare for future needs while making IT investments, which hindered the ability to extract knowledge from legacy systems. Headd states, "Hopefully these lessons will allow us to make smarter investment decisions in the future."

To work around old legacy systems, Headd and his team are re-engineering infrastructure to rethink and redefine the way the city shares data between agencies and departments. In doing so, Philadelphia can more efficiently use data to inform operational decisions and focus on cloud computing and data storage offerings that easily scale to meet demands. With increasing volumes of data, Philadelphia has learned that the ability to scale services is essential to meeting the needs of government agencies.

Yet a scalable infrastructure is only half the equation. Agencies must also focus on shareable infrastructures, thinking of new and innovative ways agencies can access data and information across the entire enterprise. This will help to avoid the fragmented nature of data, siloes and facilitate improved information sharing. Headd states, "Philadelphia does not have data that is just big, but we have big problems. The key to getting at those problems will be creating a shareable infrastructure." Leveraging big data will continue to be essential in facilitating improved decision making for the public sector.

Headd's insights provide a solid foundation for agencies to build their big data initiatives. As the public sector continues to develop more kinds of data, and the need for thorough data analysis continues to grow in importance, identifying strategies and best practices for big data programs is essential. Philadelphia is one of many cities starting to leverage data in new and emerging ways, which is transforming the way government operates and modernizing agencies.

5 STEPS TO LEVERAGE BIG DATA IN YOUR AGENCY

EXPERT INSIGHTS WITH SHAWN KINGSBERRY, CHIEF INFORMATION OFFICER, RECOVERY ACCOUNTABILITY AND TRANSPARENCY BOARD (RATB)

In an interview for this report, Shawn Kingsberry, Chief Information Officer, Recovery Accountability and Transparency Board (RATB), highlighted some of the innovative practices he is leading surrounding big data and emerging technology. Kingsberry's comments highlight the importance of leveraging big data in government, and how big data can facilitate new insights and streamline efficiencies for federal agencies. Currently, agencies are looking to data as a means to discover innovative solutions to complex public sector challenges. This section will highlight Kingsberry's work at the RATB, and five steps to leverage big data within a federal agency.

The Recovery Accountability and Transparency Board (RATB) is an independent agency within the federal government. This places the RATB in a unique position of being a government agency that is independent of political agenda and partisan politics. RATB's mission is to provide extensive transparency insights to federal programs as well as prevent waste, fraud and abuse related to how Recovery Act funds are spent. Recovery.gov, one of RATB's transparency websites, provide the opportunity for the public to explore recovery funds within their community. Citizens can visit the site to look at maps that provide an overview of where

Recovery Funds are going, or to research the way funds are used by entering their zip code. Additionally, constituents can even report instances of fraud, waste, apply for grants, and even share the raw or visualized data, which creates improved transparency and involvement in government activities by citizens.

Big data is more than the collection and storage of extensive volumes of data. Agencies need to leverage data to improve services and to do so, data must be organized and structured so information can be digested and used to reach agency goals. However, with the lack of standardization of data between agencies and the absence of established data communications between agencies, sharing and using data to optimize transparency and agency objectives across government has become a daunting task for organizations. "From the federal standpoint, we have a lot of work to do," states Kingsberry, especially as the amount of data developed every month has grown exponentially and is predicted to keep growing at an astounding rate. Kingsberry states:

"When you look at the expected amount of data that's going to be transferred per month externally across the web, the expectation is that by 2021 we're going to be looking at 180,000 petabytes per month.

If you go all the way back to the other side of the spectrum, in 1990 we were under 20,000 petabytes per month. That gives you the growth of technologies that have been coming online that enable customers, citizens, and government to be more connected,"

With increasing volumes of data, government agencies, such as the Recovery Accountability and Transparency Board, must put a focus on understanding the data and ensuring accuracy. This also calls attention to the implications for addressing changes in public outreach; Kingsberry states, "The expansive amount of growth is telling you that new customers and new users, are going to have expectations that when they ask questions they will get performance and integrity from the data." Therefore, improved performance and service delivery can only happen if data is accurate and timely.

Big data, when used in a quickly assembled government strategy, can become cumbersome and leave employees and constituents more overwhelmed and confused. However, when big data initiatives are thought out and developed out of concise agency objectives, big data has the potential to lead to more successful agency activities, and a more satisfied constituency. Below are fives ways, as indicated by Kingsberry, that an agency can create a successful big data program.

1. IDENTIFY A CLEAR BUSINESS NEED OR PROBLEM

Big data is an excellent tool to use in order to better inform business decisions. To fully leverage a big data initiative and identify new insights through big data programs, agency leaders cannot just haphazardly place big data into a business strategy without first identifying how data can be properly used. "The first thing is for federal government to get their head around business requirements and not get caught up in the buzzwords and the technology of what seems interesting right now," warns Kingsberry. With the rapid dissemination of information on the Internet, buzzwords can lead to hype and premature dedication to a specific technology program. Not every new computing advancement or mobile strategy will fit every agency. Federal agencies must understand their business objectives and how data can be part of the solution, prior to committing time and money into a big data project or program.

2. IDENTIFY RISKS AND ISSUES

Working with raw data and large sets of data can be challenging in government. Agencies are often working with information that should remain private within the agency and not shared externally. It is important to recognize the nature of the data before implementing a strategy. However, sometimes the issue is of an opposite nature. Government agencies can often be isolated from each other and have difficulty sharing information with other agencies. "If one of your risks is 'we need to do a better job providing more quality information or a more useful situation to our customers,' and to get there requires you to connect with other data sources that are outside of your agency, you have another issue," says Kingsberry. To obtain the best information to make the most informed decisions, agencies need to connect and share data sets, especially in an agency such as the Recovery Accountability and Transparency Board where data needs to be collected from every recovery aid-related agency. Once the issue is properly identified federal data sharing will have, according to Kingsberry, "a lot of opportunity for maturity."

Another issue that agencies must be prepared to face is realizing that poor collaboration and communication within an agency may also be a challenge in addition to implementing new technology. "Technology is not the challenge. If you get the right technical people sitting at the table, you can solve any technology problem. The challenge is people and culture," states Kingsberry. Departments have their own culture and way of working through issues, but those cultures may clash. To avoid conflicting ideas, Kingsberry suggests that agencies "Remove technology from the conversation. Talk at the business level about the outcome. With the

FEDERAL DATA GROWTH: BY THE NUMBERS

1990 20,000 petabytes / month

2021 180,000 petabytes / month

right people, the business side and the technology side of the organization will be sitting at the table together. What you're going to see is that you are going to begin to get people on the same page."

3. CRAFT A CLEAR VISION

Good leaders not only craft a clear vision, but also know the attributes, quality, and integrity of data to set realistic expectations. To better understand data, Kingsberry suggests that agencies ask questions such as "What is the data? What are its attributes? What is the level of quality of the data? This assures that I have an appropriate level of data integrity." These questions can help agency leaders know not only the limitations of the data, but also the potential of the data. They can then understand how to leverage data in the best way to accommodate the growing expectations of customers and to fit into a plan designed by leaders of the agency. Recognizing the scope will allow agencies to properly leverage data to accomplish the agency's mission objectives.

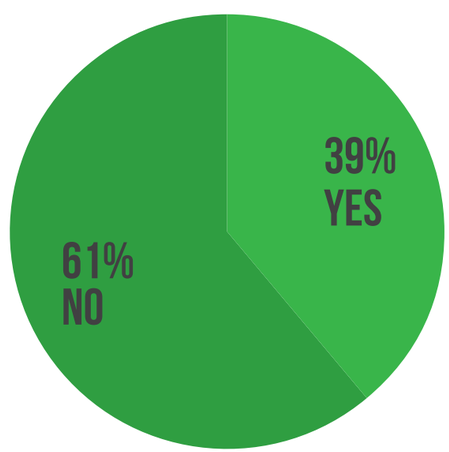
4. IDENTIFY RESOURCE NEEDS

While big data is extremely helpful when leveraged to make decisions, agencies need to understand the resources that they have and extract actionable insights from data. Kingsberry advises when implementing a new technology, whether it is a mobile strategy or utilizing the cloud, agencies need to look and understand their portfolios. A portfolio is a list of technologies and their appropriations, laying out the limitations of technological and fiscal resources of an agency. "In essence, it is all about project management and portfolio management," says Kingsberry. He cautions, "Agencies don't have to think differently," meaning analyzing and understanding your resource needs is not just about thinking innovatively outside of the box. Instead, agencies must identify objectives and recognize the work that must be done in order to accomplish their goals. To do this, Kingsberry suggests asking certain questions about your portfolio:

- What are agency risks and issues associated with delivering the goals and objectives to meet the vision of my portfolio?
- How can technology better optimize the outcome, and better meet goals and objectives of the agency?

FROM THE GOVLOOP SURVEY: DOES YOUR ORGANIZATION HAVE DATA SCIENTISTS ON STAFF? (FIGURE 6)

As education was identified as one of the main challenges for big data initiatives, 61% of agencies indicated they do not have data scientists on staff. This information is further evidence that an essential part to adopting big data will mean providing new training opportunities for existing employees, and leveraging their skill sets in new ways. Although the title "data scientist" may not be a formal role within an agency, leaders must certainly continue to train employees to serve in a data scientist capacity. This means leaders must have employees on staff that focus on data, and learn the best ways to identify intelligence from data collected across the agency or department, and collaborate with partnering agencies.



5. ENABLE AND OPTIMIZE SERVICES

Kingsberry pointed out that very recently, the RATB was able to "run two million records against 100 million rows in roughly two minutes," a process that used to take hours, if not days, to run. "We're constantly looking to optimize," states Kingsberry. The ability to quickly and efficiently find knowledge out of data means that data can become more actionable, and lead to stronger insights for improved decision-making. For instance, at the RATB, the ability to quickly process millions of records leads to improved decisions around Recovery Funds and transparency initiatives.

The purpose of the Recovery Accountability and Transparency Board is to provide information to the public about Recovery funds so that other agencies, businesses and individuals can see where government money is going. Publishing a report or posting information on a website is helpful in reaching out to constituents but in order to optimize activities and transparency, agencies need to consider mobile technology. At RATB, "customers can access our infrastructure via tablets, because of the infrastructure when you look at our analytic cloud. So in essence, that is mobile and mobile is a way of interacting with the data," says Kingsberry.

As previously stated, big data is more than just acquiring immense amounts of information, big data is about actionable insights and is an important part of a digital government strategy. "You can't talk big data without talking mobile, without talking cloud, and without talking information assurance. They all go hand in hand because at the end of the day it's all about the data. Everything else is how you interact with that data." If agencies leverage the data sets that they acquire and utilize technologies such as mobile and cloud, then they will be able to better serve their constituents and meet agency objectives. The important action to take, as Kingsberry has expressed, is to first identify the scope of the data and to fit that into the mission of an agency. Once that is accomplished, big data can aid government agencies in improving transparency and public services.

CASE STUDY: ACHIEVING WORKFORCE EFFICIENCIES WITH BIG DATA

An interview with industry expert, Audie Hittle, Chief Technology Officer (CTO) Federal Market at EMC Isilon

Recently, Audie Hittle, Chief Technology Officer (CTO), Federal Market at EMC Isilon, shared his insights as to how big data initiatives are transforming government and the role of government employees. As Hittle states, "We are only at the tip of the iceberg and I believe big data can be transformational." Hittle's statement could not be more poignant, as agencies are looking to streamline processes and identify new efficiencies.

As this report has shown, big data analysis has many benefits for government agencies; "One of the areas big data certainly offers is the opportunity to make better decisions and make more efficient use of all resources," Hittle states. Across all levels of government, focusing on finding new efficiencies is essential, especially in a time where budgets are being reduced, coupled with high levels of fiscal uncertainty.

Yet, efficiency for government does not just mean the benefits of quicker processing, more efficient services, and improved decision making. This also means the improved efficiencies within the job of a government worker. Hittle notes, efficiency "enables people to get more gratification out of their job, because they feel they are more efficient, they are accomplishing more, they are contributing more, and those are great things."

Hittle shared a case study from the Air Force, which identified the improved efficiencies for the workforce. The Air Force implemented a big data program and was able to reduce the number of dedicated storage professionals. In doing so, the Air Force was able to reallocate staff and place employees on more mission critical and support roles. Hittle states, "in some cases you can see where placing employees on mission critical roles provides a career progression or a career re-training opportunity to move from one highly skilled area, where it deals with IT management, into other areas where it deals with knowledge extraction or analysis."

The Air Force is not alone in making improvements to the workforce. A second case study comes from an intelligence agency, which implemented an Isilon scale-out network-attached storage (NAS) solution. With this solution, the agency was again able to reduce service professionals by roughly 90% and, similar to the Air Force, employees could be reallocated to higher priority and direct mission-oriented services.

In addition to transforming how the workforce operates, Hittle believes that big data will also facilitate new opportunities for government training. "What we are seeing here is that big data is going to create efficiencies that are going to lead to cross training and retraining opportunities that enable people to migrate quickly into these high priority roles," states Hittle.

Clearly, big data can help agencies transform their decision-making and service delivery. The impact on the government workforce cannot be overlooked, with tightening resources, IT managers must assure they are fully leveraging emerging technology as a way to redefine the role of a government employee, and fully leverage their skills and knowledge. With new efficiencies comes the ability to free workers from monotonous, time-intensive tasks, and allows workers to use their skills to contribute to higher value mission-centric objectives.

HOW TO IDENTIFY ACTIONABLE INSIGHTS FROM BIG DATA

An Interview with Industry Expert Pat Herbert, Principal Solutions Architect, SAS

Pat Herbert, Principal Solutions Architect, SAS, recently spoke with GovLoop on the importance of big data, common challenges and how agencies can leverage big data to improve internal operations at a government agency. Herbert's insights show how agencies can begin to find actionable intelligence from big data, and leverage their data in new, innovative ways.

Like all our expert interviews in this report, Herbert cautions agencies not to get lost in defining big data in terms of size or scale. Herbert states, "We do not define big data as a function of scale or size exclusively. Big data is the amount of data or the complexity of data that exceed a given users comfort, their ability to handle that data in ways they have traditionally handled it, so it is something that's bigger than they have dealt with in the past but bigger does not always mean size"

For Herbert, the ultimate goal is taking big data and finding new value, actionable insights and using data as a strategic tool to transform an agency. During the interview, Herbert mentioned some common challenges and best practices for big data analysis. One of the first challenges that Herbert mentions is the challenges related to governance, "Ultimately a challenge is a data governance issue, the ability to use the data the way agencies need to in order to access, but also in a way that is effective for agencies."

Governance challenges are also multiplied as data is often fragmented across a variety of different agencies. Along with governance challenges, this also results in challenges related to collaboration and extracting value from data. "The challenge is not having data, it is having access to data and then being able to use it intelligently," Herbert states.

The goal with big data analysis is to find new associations between variables, that otherwise would remain unknown. An example would be to look at the data collected by a school district, transportation department and public safety data, which all may have data points that looked at in isolation may seem irrelevant, but when merged with other variables, new associations and trends can be found.

Yet, the ability to share data across departments and agencies remains a challenge at all levels of government. Herbert states, "Data infrastructures aren't typically put together in a way that permits them to easily share data, and as a result there is the potential to be able to have extremely valuable data, but their execution ends up locking them out of exercising their full potential from that data."

One of the lessons learned that Herbert shared with GovLoop is the necessity to focus on balance while implementing a big data program. Agencies must approach big data with a clear scope, defined goals and defining what success looks like. Herbert says he has seen both sides of the spectrum in adoption, such as agencies looking to adopt an enterprise wide strategy right off the bat, and also agencies looking to adopt smaller, less mission-focused programs.

The key for agencies looking to leverage big data is finding the right balance and knowing the appropriate level of investment in a big data program. Herbert states, "Identify a project that may not be the most critical, but is an important project. Work on that as a complete initiative, and then expand on that rather than trying to build out a big data initiative that covers every single bit of data." In doing so, agencies can begin to unlock actionable intelligence from big data.

USING BIG DATA TO PREVENT WASTE, FRAUD AND ABUSE

To witness improved cost savings in government, there has been an increased interest in using big data analysis to detect fraud, waste and abuse. Upon analyzing various data, the Government Accountability Office (GAO) realized that fragmentation, overlap and duplication were an issue in government, causing an immense waste of funds. Nikki Clowers, Director of Financial Markets and Community Investment Issues at GAO, explained in an interview with GovLoop's DorobekInsider the nature of fragmentation.

"Fragmentation refers to a situation when you have multiple agencies providing services in the same broad area. If there is a big issue facing the nation you can imagine there are a wide number of agencies that might be involved; but what we find is, fragmentation can be a harbinger for overlap and duplication which is a more serious issue." Fragmentation can greatly affect the efficiency of agency programs. Often, information can fall through the cracks of several programs because of miscommunication on which agency is responsible for what aspects of an issue.

Fragmentation often leads to duplication and overlap. Duplication, according to Clowers, is when "the same agencies in the same broad area provide the same services to the same beneficiaries." This is different from overlap, which is when "you could have multiple agencies in the same area with similar activities but the beneficiaries might be a little different,"

Over the years, GAO has researched where duplication and overlap have occurred in order to identify those instances in order to prevent them.

states Clowers. In the end, both fragmentation and duplication lead to a significant waste of government resources.

Over the years, GAO has researched where duplication and overlap have occurred in order to identify those instances in order to prevent them. Clowers says, "We [GAO] start with the budget, and looking at different functions at the highest level, you can see where there could be a potential for overlap. From there we drill down to the strategic plan and other agency documents. We go out and interview officials. We go on-site to see the services and activities being performed." More often than not, the cause of the redundancy is missing data, whether it is lack of budgetary information or performance data. That said, it is important to note that redundancy can be useful for agencies, such as Department of Homeland Security and law enforcement. However, the overlaps must be coordinated, "because if it is not properly coordinated, that is where we will get confusion and gaps," said Clowers. The take-away from Clowers and the work at GAO, in the simplest terms, is that big data matters.

Additionally, Jennifer Kerber, President of the TechAmerica Foundation, spoke on GovLoop's DorobekInsider, providing insights to a recent TechAmerica survey that confirmed the importance of big data. "Budgets are tight especially with the government trying to cut $85 billion in sequestration cuts. So I think there is a significant focus on how big data can help reduce waste, fraud and abuse," states Kerber.

As previously mentioned, acquiring and using data is one way to identify and prevent waste. However, disorganized data might be more detrimental than no data at all. In an interview with Kerber, she detailed several three types of data that could help in saving money.

1. CONTROLLED DATA

In today's technologically based world, data is developed and collected at incredible rates. While many agencies are able to collect this data, often the volume can become cumbersome and, therefore, the data remains unorganized. "We are just now at a point where the amount of data that's coming in and the ability to control that data is allowing us to have 'Smart Data,'" says Kerber. 'Smart Data,' or data that is organized into sets, can prove to be an excellent resource for running analytics software to detect waste or fraud.

2. REAL-TIME DATA

"The ability to do some of this analysis in real-time allows you to better deliver services to customers," states Kerber, "Public safety officers can now use big data to reduce crime and target criminal hot-spots. One of the most exciting areas is health care. You can do targeted searches in cancer research for example." Data that is no longer relevant can be dangerous when agencies depend on it to plan actions or programs. Accurate data that is presented in real-time allows agencies to make quick decisions based on updated information.

3. SYNTHESIZED DATA

No matter how much data is collected, it is not very useful if data sets cannot communicate with each other. "Data silos are also a big concern," says Kerber. It's important that data sets from different parts of an agency are interconnected so there is not duplication within an agency. That said, it is also vital that data from various agencies across government can be synthesized in order to detect overlap and waste on a larger scale. These efforts are still being developed, according to Kerber: "Right now a bunch of smart people are working to make all the data able to talk to each other."

With the government facing tight budgets while promising to provide transparency, it cannot be denied that big data is important. "The potential is enormous," Says Kerber, "It's evolutionary. It's almost like the Internet. These days nobody asks if you are on the Internet anymore, it's just a way of life. Eventually we see that same evolution in big data. The data analytics and the decisions behind it will just become a part of life." The key is to focus not just on collecting and using data, but harnessing the data to turn it into organized and easily usable information that can inform agency decisions and save hundreds of millions of dollars.

DOES CONGRESS MANDATE DUPLICATION?

Sometimes Congressional legislation causes duplication. One such instance is with catfish regulation. "In 2008," says Clowers, "the Farm Bill gave the responsibility for inspecting catfish to the USDA, but there were already two other agencies performing that function. We (the GAO) have recommended that the Congress take those responsibilities away. It is costing the USDA $14 million annually to regulate the catfish."

YOUR BIG DATA CHEAT SHEET

This report provided you with many lessons learned, best practices and case studies from thought leaders in government and industry who are leveraging big data to transform government. The following sections provide you with your big data cheat sheet, a brief summary of the report and resources for you to find more information on big data. Big data analysis has the potential to radically transform how government operates. As we continue to witness an explosion of data, increasingly agencies need to focus on practices and strategies to extract knowledge and actionable insights.

ALREADY KNOW THE NUTS AND BOLTS OF BIG DATA? LOOKING TO GET SMART QUICK ON BIG DATA?

In either case, here is a quick reference sheet for you to review the need to know information on big data.

WHY SHOULD GOVERNMENT LEVERAGE BIG DATA?

There are lots of benefits, but here are the top 5 benefits to remember with big data:

1. Improve efficiencies

2. Improve services

3. Reduce costs

4. Allocate resources more effectively (do more with less)

5. Reveal new insights that otherwise remain unknown

WHAT DO I NEED TO CONSIDER?

There are lots of elements to consider and, of course, this cheat sheet can't get to them all, but here is the need-to-know information to get a strategic conversation started:

1. Who knows (or can learn) big data analysis on staff? Do we have skills in house?

2. How do we collaborate and what are the core privacy and security issues?

3. Who creates data and what kind of format is it in?

4. What are we already doing to extract value from our data and how can we improve?

5. What specific problem are we attempting to solve?

WHAT ARE SOME EXAMPLES AND RESOURCES TO REFERENCE?

1. Case Study: Boston - Street Bump

2. Case Study: New York City - The Mayor's Greek Squad

3. Case Study: Big Data for Cancer Care

4. Article: Big Data Revolution, Changing the Way We Live, Work and Think

5. Article: Big data: A Big Problem that's Getting Bigger

KEY BIG DATA TERMS AND GLOSSARY

HADOOP

Hadoop is free to use and is a Java-based framework that supports processing large data sets that are from disparate servers and hosted in various locations.

MACHINE LEARNING

This is a type of artificial intelligence that allows computers to learn without being programmed – computers can adapt and grow based on data.

NOSQL DATABASES

"Not Only SQL," address challenges for big data that relational databases cannot. It's an approach to database design for large sets of distributed data, useful for data stored in the cloud, virtual and unstructured.

NATURAL LANGUAGE PROCESSING (NLP)

Component of artificial intelligence, allows computers to understand human speech.

MAPREDUCE

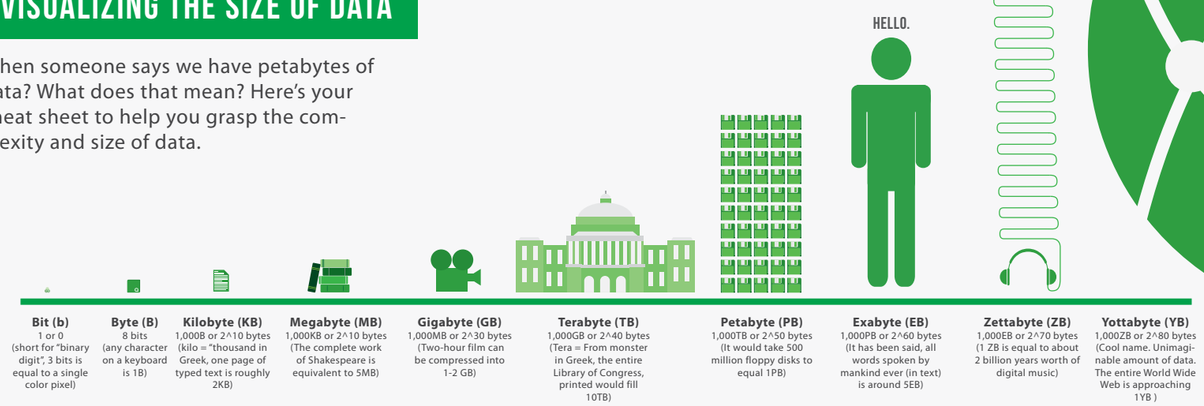
A two-part process to understand massive amounts of unstructured data; map: parcels out data, and reduce: merges data together, allows programmers to run programs across multiple machines.

For more definitions, check out some of the following:

- Big Data, Bigger Opportunities: Investing in Information and Analytics (Gartner Report)
- Big Data Glossary (O'Reilly Media)

VISUALIZING THE SIZE OF DATA

When someone says we have petabytes of data? What does that mean? Here's your cheat sheet to help you grasp the complexity and size of data.



4 CORE STEPS FOR BIG DATA

Here are four high-level steps for you to always consider for implementing big data programs:

1. Learn from others that are leveraging big data successfully
2. Build a strategic framework for harnessing the potential of big data
3. Understand the human resource impact
4. Determine the types of data available to you

5 NEXT STEPS & CONVERSATION STARTERS

This guide provided you with some best practices, case studies and expert insights to help you fully leverage big data. But, we know that you're busy – so just like we gave you a cheat sheet for big data, here's everything you need to know about this guide, based on different settings you may find yourself in.

1. THE TWITTER VERSION

Reading @GovLoop's latest report on big data – handful of case studies and best practices: bit.ly/117vrGE

2. THE FACEBOOK VERSION

Check out GovLoop's report on big data – cheat sheet on big data, dozens of case studies – easy way to get smart on big data. http://www.govloop.com/profiles/blogs/transforming-your-agency-with-big-data

3. THE LINKEDIN VERSION

Check out GovLoop's report on big data – includes big data survey from GovLoop community and case studies sharing best practices, lessons learned and overcoming common challenges for big data initiatives. http://www.govloop.com/profiles/blogs/transforming-your-agency-with-big-data

4. HAPPY HOUR SMALL TALK BULLET POINTS

- The data government collects has the power to change the way government does business, but government is challenged to extract knowledge and actionable intelligence.
- Philadelphia has implemented dozens of open data programs including releasing an Open Data Guidebook.
- Mayor Nutter named Mark Headd as Philadelphia's first Chief Data Officer, Headd is responsible for overseeing open data and transparency initiatives across Philadelphia.
- The Recovery Accountability Transparency Board (RATB) is leading many big data programs in the federal government. They worked with dozens of agencies to provide insights on the how stimulus money has been used to develop Recovery.gov.
- IDC has stated that the big data market is expected to grow from $3.2 billion in 2010 to $16.9 billion in 2015 (Source)
- You need special technology to do this (MapReduce/Hadoop) – but can't forget about leadership, vision, culture and workforce needs
- Eric Schmidt stated in 2010 that every two days we create as much as information as we did throughout the entire history of civilization up until 2003 (Source).

5. THE BIG DATA 30-SECOND ELEVATOR PITCH TO YOUR BOSS AND PEERS

Have you had a chance to take a look at GovLoop's Big Data report? They have a handful of case studies and best practices, and I think I got some ideas on how we can start to leverage our data. Can we spend a half hour to talk through? Would love to float some ideas and see if they can work for our agency.

Big data is changing the way government does business. Big data is typically defined with the three V's: volume, variety and velocity. As agencies are now collecting data from a variety of sources (social media, sensors, CRM, and ERP data), agencies need new and innovative ways to find insights and leverage the value of the data they are collecting. There are countless examples (see above to fill your pitch with some great case studies) of leveraging big data to improve safety, public health and customer service in government.

ACKNOWLEDGEMENTS

The GovLoop team is thankful to all of those who contributed to this report. We thank everyone for their participation in our survey, active community engagement, input and knowledge shared while developing this report. Thank you to Bob Gourley for writing the foreward of this report.

This guide would not have been possibility without your assistance and from the support of our sponsors, EMC and SAS.

Lead Author:
PATRICK FIORENZA, *Senior Research Analyst*

Co-Author:
KATE LONG, *GovLoop Spring 2013 Fellow*

Editor:
ANDREW KRZMARICK, *Director of Community Engagement*

Editor:
STEVE RESSLER, *GovLoop Founder and President*

Lead Designer:
JEFF RIBEIRA, *Senior Interactive Designer*

Designer:
RUSSELL YERKES III, *GovLoop Summer 2013 Design Fellow*

For more information about this report, please contact Patrick Fiorenza, at pat@govloop.com.

ABOUT GOVLOOP

LOCATION

GovLoop is headquartered in Washington D.C., with a team of dedicated professionals who share a commitment to connect and improve government.

GOVLOOP

734 15th St NW, Suite 500
Washington, DC 20005
Phone: (202) 407-7421
Fax: (202) 407-7501

GovLoop's mission is to connect government to improve government. We aim to inspire public sector professionals by acting as the knowledge network for government. The GovLoop community has over 65,000 members working to foster collaboration, solve problems and share resources across government.

The GovLoop community has been widely recognized across multiple sectors. GovLoop members come from across the public sector. Our membership includes federal, state, and local public servants, industry experts and professionals grounded in academic research. Today, GovLoop is the leading site for addressing public sector issues.

GovLoop works with top industry partners to provide resources and tools to the government community. GovLoop has developed a variety of guides, infographics, online training and educational events, all to help public sector professionals become more efficient Civil Servants.

GovLoop's report, Transforming Your Agency with Big Data is sponsored by EMC and SAS.

If you have questions on this report, please feel free to reach out to Patrick Fiorenza, Senior Research Analyst at pat@govloop.com.

RESOURCES

Big Data Resources for Public Sector Professionals
http://www.govloop.com/big-data

Are Big Data and Cybersecurity a Perfect Match? Emily Jarvis, GovLoop
http://www.govloop.com/profiles/blogs/are-big-data-and-cybersecurity-a-perfect-match

Turning Data in Decisions – Why Metrics Matter. Emily Jarvis, GovLoop
http://www.govloop.com/profiles/blogs/turning-data-into-decisions-why-performance-metrics-matter

Big Data, Big Fear, Big Potential. Danielle Blumenthal
http://www.govloop.com/profiles/blogs/big-data-big-fear-big-potential

Big Data, Social Media and the Long Tail of Public Policy. Nicholas Charney, Posted on GovLoop
http://www.govloop.com/profiles/blogs/big-data-social-media-and-the-long-tail-of-public-policy

Learning from the World Bank's "Big Data" Exploration Weekend. Dennis McDonald, posted on GovLoop
http://www.govloop.com/profiles/blogs/learning-from-the-world-bank-s-big-data-exploration-weekend

Transparency Report Card - Did Your State Make the Grade? Emily Jarvis, GovLoop
http://www.govloop.com/profiles/blogs/report-cards-are-in-did-you-state-s-transparency-make-the-grade

Data as Paint, and the Rise of the Data Artist. Bob Gourley, posted on GovLoop
http://www.govloop.com/profiles/blogs/data-as-paint-and-the-rise-of-the-data-artist

Arlington Cemetery Goes Mobile - Take a Look at the New ANC Explorer App. Emily Jarvis, GovLoop
http://www.govloop.com/profiles/blogs/arlington-national-cemetery-goes-digital-take-a-look-at-the-new-a

How Big Data and Location Analytics are Impacting GIS for Government. Pat Fiorenza, GovLoop
http://www.govloop.com/profiles/blogs/how-big-data-and-location-analytics-are-impacting-gis-for-governm

Government Big Data: What's Next? Experts Give Insight. Bryce Bender, GovLoop
http://www.govloop.com/profiles/blogs/government-big-data-what-s-next-experts-give-insight

Obama Administration Unveils "Big Data" Initiative: Announces $200 Million in New R&D Investments. Press Release, Office of Science and Technology Policy, Executive Office of the President
http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

734 15th St NW, Suite 500
Washington, DC 20005
Phone: (202) 407-7421
Fax: (202) 407-7501