

HOW YOU CAN USE

DATA ANALYTICS

TO CHANGE

GOVERNMENT

CONTENTS

2 | What Is Data Analytics?

4 | Fighting the Opioid Epidemic



7 | How to Make Data Analytics User-Friendly

8 | Making Patent Data Searchable



11 | How Data Analytics Drives Real-Time and Forensic Analyses

12 | Tracking Medicare Fraud



15 | Accelerating Insights with Multi-Tiered Flash Storage

16 | Improving Human Capital Management



19 | The Future of Data Analytics in Government IT

20 | Empowering Cyber Analysts



23 | In the World of Data Analytics, Speed Is Everything

24 | Predicting and Reducing Crime



27 | 4 Keys to a Successful Data Analytics Program

28 | Improving Mail Delivery



31 | Data Analytics as an Asset

32 | Enhancing IT Investment Tracking



34 | 3 Reasons Privacy Should Be Top of Mind

35 | What You Can Do Now: How to Create a Culture for Data Analytics

37 | About Govloop & Acknowledgments

EXECUTIVE SUMMARY

Everything about Dr. Farid Fata's case seemed too egregious to be true.

When a [whistleblower tipped off the Department of Justice back in 2013](#) about Dr. Fata's unprecedented medical malpractice, government officials had a hard time wrapping their heads around the claims — at first.

How could a renowned Michigan oncologist intentionally misdiagnose patients and then inject them with chemotherapy treatments? And how could he intentionally underdose patients who did in fact have cancer?

In total, 553 patients and their loved ones were victimized. Some patients died before seeing Dr. Fata sentenced to 45 years in prison on July 10, 2015. Others were left with organ damage, missing teeth, nerve damage and psychological and emotional scars.

Because of the nature of the harm, government officials are calling Dr. Fata's case the most serious fraud case in U.S. history. Bringing him to justice took a brave whistleblower, government collaboration — and reliable data analytics.

"Our investigators partnered with the FBI, IRS and DoJ, including the U.S. Attorney's Office in the Eastern District Michigan," said Caryl Brzymialkiewicz, Assistant Inspector General and Chief Data Officer at the Department of Health and Human Services Office of the Inspector General. "They all worked collabora-

tively, nonstop, and pulled enough evidence together to take quick action, including the use of data analytics."

When this case came in, the team checked the whistleblower's lead against different data sets, including claims data, Brzymialkiewicz said. The data corroborated the whistleblower's claims, but initially it was such an outlier that officials thought something was wrong with the data.

"We talk about two outcomes for this office: It's really about protecting beneficiaries, as well as the programs," she said. "So I tell that story because, in terms of detecting fraud, using data analytics is a huge mission."

Healthcare is one of the key areas in government where data analytics is driving value, but it's certainly not the only one. In Indiana, government employees are using data analytics to combat the state's opioid epidemic. The General Services Administration is using human capital analytics to inform human resources hiring and recruiting decisions. And there are countless other examples, several of which you'll find in this guide.

When used effectively, data analytics can help to save lives, improve efficiencies, reduce costs, and help government deliver better citizen services. And government is using these capabilities more than ever, so it's especially important today for you to learn how and why they function.

To support you on your data analytics journey, we've put together this GovLoop resource to help you:

- **Define data analytics and understand why it's a big deal in government**
- **Identify various use cases at the federal, state and local levels and their underlying key success factors**
- **Understand the privacy concerns associated with data analytics**
- **Glean practical tips for building a culture for data analytics at your agency**

But before we dive into the meat of this guide, let's establish the basics about data analytics — what it is, and why you should care.

WHAT IS DATA ANALYTICS?

Is it big data, data analytics, or just analytics?

If you've ever pondered this question, you're not alone. Often you'll hear these words used interchangeably or merged into one phrase: big data analytics. It can be confusing, right?

But don't get too caught up in the terminology because it's bound to change from person to person. For the sake of this guide, we will stick with a definition provided in President Obama's [Big Data and Privacy Working Group Review](#):

“**Analytics is what makes big data come alive.**

Without analytics, big datasets could be stored, and they could be retrieved, wholly or selectively. But what comes out would be exactly what went in. Analytics, comprising a number of different computational technologies, is what fuels the big-data revolution. Analytics is what creates the new value in big datasets, vastly more than the sum of the values of the parts.”

One of the reasons you'll often hear big data, data analytics and big data analytics used in the same breath is because there is a lot of overlap. We need both the tools to collect and manage large volumes of varying data, as well as the tools to make sense of the data and turn that insight into action.

When you hear the word “big” coupled with “analytics,” consider these two things: If the data that's collected and processed is big in quantity and variety, then it's likely the analysis and data correlations will be big too.

But data analytics is about more than algorithms. It's also about the physical platforms agencies need to store and analyze their data. To meet those needs, many agencies are turning to cloud computing, while also balancing the need for security and privacy. We'll get into the major privacy issues later in the guide.

4 TYPES OF DATA ANALYTICS

Until this point, we've used data analytics as a blanket term to describe how agencies glean insight from their data to solve real problems and carry out their missions. But there are four distinct types of data analytics that help agencies answer the following questions ▶

We've pulled together some insights from Gartner to help you understand the four types of analytics and visualize the rising degree of difficulty and value tied to each of them. You'll notice these terms (descriptive, diagnostic, predictive, and prescriptive) mentioned throughout the guide, especially in our case studies.

- 1 What happened?
- 2 Why did it happen?
- 3 What will happen?
- 4 How can we make it happen?

ANALYTIC VALUE ESCALATOR



APPLY IT NOW

Take a minute and think about one of your organization's biggest pain points, or perhaps a challenge you face in your everyday work life. Maybe there's a specific project you're working on right now. Think about the type of insights or efficiencies you could gain if you understood what happened, why it happened, what will likely happen, and what to do about it.



The State of Indiana



An interview with Dave Matusoff & Joshua Martin

Director for the Indiana Management & Performance Hub; MPH Chief of Staff

THE CHALLENGE: Fighting the Opioid Epidemic

In Indiana, news stories about Hoosiers who are addicted to opioid are driving more open conversation about the state's growing drug epidemic. The stories are about people like Cheryl Dicken, who, [according to the Indianapolis Star](#), "found her son unconscious in a heroin-induced haze," and Joan Moon, "whose grandson has been addicted to drugs since 2009."

These hurting families are the face of a larger problem that demands more answers, more insight and, inevitably, more meaningful data. Although personal stories humanize the drug epidemic, stories alone are not the solution. Law enforcement officials, public health agencies, lawmakers, and the community need sound data to shape their collective response to the problem.

But when that data is stored in different systems at different agencies, making use of it requires hard work and collaboration.

Spurred by serious concerns about the rise of opioid use statewide — and a fascination with data — Joshua Martin, Chief of Staff for the Indiana Management and Performance Hub (MPH), decided to act.

It was in August 2014 that Martin noticed a rise in media coverage of heroin-related incidents. That was about the same time MPH deployed a new business intelligence tool to help its agency customers make data-informed decisions.

The culmination of these events got Martin thinking: Wouldn't it be interesting if the state could use data to better understand the changing nature of the drug problem?

Historically, decision-makers have had to rely on whatever information was available to make the tough choices, but what they had access to was not always the best information. Sometimes the people who were asking the questions didn't have access to the right data, and neither did the people writing policies.

"There's a real human element to using the data to inform the policy decision-making around these problems," said Dave Matusoff, Executive Director of MPH.

THE SOLUTION

“What we’re trying to do here is take these mountains of data that are in different places and turn data into information, turn that information into knowledge, and hopefully turn that knowledge into policy.”

It has only been a few months since MPH stood up its drug data dashboard internally and began enlisting the help of other agencies — the data owners. The project is still in its infancy, but that hasn’t detracted from the real work underway.

The MPH team has worked closely with Indiana’s police forensic laboratories to track and visualize the types of illegal substances that officers are finding on the street, and how often they’re finding them. The most recent data (to November 2015) showed a dramatic rise in heroin submissions to the lab but a steady decline in other opioid submissions, namely prescription drugs.

Matusoff pointed to a [2013 state law](#) aimed at tackling the opioid epidemic as a contributing factor in the decline of prescription opioid abuse. The law was passed to tighten prescribing practices, close down so-called “pill mills” where physicians and other providers illegally prescribe and issue narcotics, and reduce opioid abuse and ultimately deaths. The General Assembly’s intentions were good, but the data shows unintended consequences.

“When they passed that legislation, it was harder for addicts to get prescription drugs,” Matusoff said. “You saw a dramatic decline in the number of prescription narcotics submitted to the forensics lab, but then you saw a really dramatic rise in the heroin use. So you can begin to look at — in a very real-time way — the unintended consequences of policy.”

In November 2015, the MPH team started formalizing the dashboard

and soliciting data and input from law enforcement and public health agencies. Using the business intelligence tool, MPH can visualize data by creating geospatial maps, color-coordinated line graphs with trend data and even a time-lapse map that shows how the drug epidemic has evolved over time.

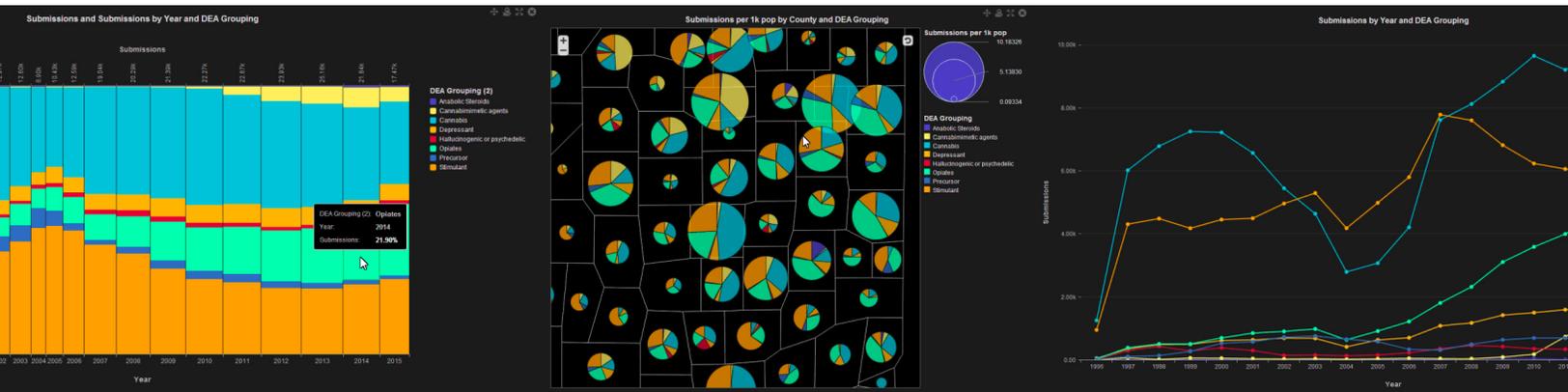
MPH isn’t just collecting data on laboratory-tested substances. The team is also looking at opioid-related deaths and geospatial data for certified treatment providers across the state.

This is the first time all these data sets have been combined to tell a complete story, Martin said. A quick look at these visual maps makes clear that the drug problems vary by county.

One color-coded map in particular shows the number of per-capita, opioid-related deaths by county. Southeastern Indiana has been hit by this epidemic especially hard. In October 2015, [the Indianapolis Star reported](#) “more than 180 people in or near Scott County were diagnosed with HIV after sharing needles to inject prescription pain medication.”

MPH is exploring the role that data will play in combating the state’s devastating drug problems. The team also has its sights set on providing similar data analytics services to other states facing their own issues.

“This is still a work in progress, but what we’re trying to do here is take these mountains of data that are in different places and turn data into information, turn that information into knowledge, and hopefully turn that knowledge into policy,” Martin said.



SUCCESS FACTORS

1
Think through your strategy. Create a proof of concept, or partner with other organizations that have already made investments.

2
Rethink how you can use your data in new and meaningful ways.

3
Solicit the help of administration officials to break down data-sharing silos among agencies.

ORACLE®

DATA ANALYTICS

LEARN MORE

<https://www.oracle.com/solutions/business-analytics/resources.html>

How to Make Data Analytics User-Friendly

An interview with Rob Lindsley

Director of Data Analytics, Oracle Public Sector

By now, data analytics has become one of the biggest trends in government. Agencies need it, companies are offering it, and solutions are driven by it. One of the most important aspects of data analytics is its ability to integrate with existing systems, in terms of data sources, applications, infrastructure, as well as security. But it can take a lot of work to reach that level of integration.

Rob Lindsley, Director of Data Analytics for Oracle Public Sector, spoke with GovLoop about ways that organizations can work to make data analytics as user-friendly and flexible as possible. Oracle's own strategy is focused on providing organizations with a variety of cloud options — allowing for self-service analytics and data visualization — and presenting big data and predictive analysis without a steep learning curve. Simply put, the strategy is to give customers “access to the data they need paired with the analytics they want,” according to Lindsley.

The way government agencies provide analytics to their users is also changing. They can run analytic solutions in their own datacenters; companies that offer the analytics services can deliver them as cloud offerings; or agencies can have third parties run their datacenters. Regardless, the tools or services should be easily available to consumers both inside and outside the organization.

Lindsley described Oracle's role in government agencies as running the gamut from core business operations like finance and human resources to mission support functionality and beyond. “Oracle has a strong presence in the central operations of many government agencies, but we also have a lot of customers using analytics to accomplish

their agency's core mission in a variety of areas, including transportation, a child services and in defense organizations. They're using analytics to meet their objectives quickly and more efficiently.”

One of the hallmarks of Oracle's analytics platform is that it's offered in both an on-premise deployment model or in the cloud. Users can select which option best suits their needs. For the cloud option, users can choose a public, private or hybrid model. Flexibility is key, because users can easily change between on-premise deployment and cloud deployment as projects develop and agencies' needs evolve over time. For example, an agency might test a new system in the cloud and then move it to an on-premises datacenter for production deployment. Or they might use cloud infrastructure for public-facing websites and offer improved data transparency.

A variety of data sources are essential to improve flexibility for analytics platforms. Traditional data analytics employs “official” data warehouses or other standard data models, but increasingly government employees need to integrate their own data into the analysis, whether through an Excel spreadsheet or a current roster of government contractors. Oracle delivers self-service capabilities that allow end-users to incorporate a variety of data sources into their analytics.

Agencies are also pushing to instill a culture of data-driven decision making. Both big data analytics and predictive analysis have significant potential for employees when they can be used without a steep learning curve. The government has vast amounts of information on the services it provides, and it

also has access to information about external events. An example of an agency successfully harnessing the power of big data analytics is the Department of Transportation (DOT) in a mid-western state. By combining data on traffic patterns, road conditions, safety features, and the weather, this agency can optimize the travel experience, make best use of limited resources and increase the safety of drivers across the state.

Oracle has also helped agencies use existing data to improve current projects and predict future trends. One project involved the U.S. Army, which for more than a decade has worked with Oracle to help identify which units are best suited for particular missions. The process aggregates numerous datasets, including those for staffing, training levels and existing resources. The Army started with straightforward metrics to evaluate unit readiness and then moved toward descriptive analysis, using information about specific capabilities unique to the unit. More recently, the Army is using predictive analysis to determine what types of units will be required in future scenarios.

Lindsley notes these examples highlight the democratization of data access, the usability of modern analytics tools, and easy implementation. Oracle's strength is solving complex technology challenges, he says. “For government, this means our analytics initiatives can help agencies address a variety of requirements – whether it's massive data sets, stringent security, or scalability. These examples remind us what can be achieved when we can seamlessly incorporate data analytics into our problem-solving strategies.”



U.S. Patent & Trademark Office



An interview with Ted Parr

Director of the Public Information Services Group within the Office of the CIO at USPTO

THE CHALLENGE: Making Patent Data Searchable

We can all agree that managing a growing database of more than 6.7 million records is no small feat. What makes this undertaking even more challenging is figuring out how best to make decades-old records electronically searchable and accessible to the public.

For the U.S. Patent and Trademark Office (USPTO), getting this right was not only an agency initiative but also a White House priority.

From the public's perspective, there's a great demand for the bulk assignment data that USPTO makes available via its [online search tool](#), appropriately named Assignment Search. "We get between 2,000 and 3,000 searches a day," said Ted Parr, Director of the Public Information Services Group at USPTO. "There are about 6,000 different users each week who use the tool."

The tool allows users to search for the "assignments," or legal paperwork, that document the transfer of a patent from a previous owner to the new owner. It could be a will or a contract that includes the name of the previous owner, the new owner, their addresses and the type of transfer being made.

"We've had the search tool on the Web for quite a while, but it was rather basic," Parr said. "It had only a few of the data fields that we have in our database open for searching, and it really was just a database lookup tool. It didn't have any kind of advanced functionality, or even what you would consider basic functionality nowadays."

For example, in the old system, a search for a common assignee could yield hundreds of pages of results and require hours of analysis. To see all the assignments from Motorola to Google using the old system, a user would have to search for Motorola as the assignor. That search alone produces 685 assignments; and each one would have to be opened to see if it was assigned to Google.

There were searches that even USPTO employees couldn't do without engaging a developer to concoct a specific query for them.

THE SOLUTION

“The net result is that customers can be very precise with their searches and can filter and analyze results in many different ways.”

Today, those searches are far less cumbersome for both USPTO employees and the public, thanks to more robust analytics.

“It’s not so much that we put in place a specific analytics capability, but we enabled it,” Parr said. “The new system uses Solr open source software to provide indexed searching rather than the slower, more limited database lookups we previously used.”

In an effort to make data more transparent to the public and easier for people to find, and to enhance innovation, facilitate technology transfers and reduce abusive patent litigation and lawsuits stemming from patent infringement, USPTO focused its attention on enhancing the search functionality of patent assignments.

“The new system allows searching by several additional data fields, including execution date, conveyance type, correspondent and assignee address, as well as allowing wild card and Boolean searches for greater precision,” Parr said. “The net result is that customers can be very precise with their searches and can filter and analyze results in many different ways.”

Remember the Google and Motorola example we gave earlier? Using the new system, users simply enter both names in their respective fields, and within a few seconds they will see the 271 assignments from Motorola to Google, rather than having to sift through more than twice that number of less precise search results.

“In addition, the new tool will have a true data export feature in its next release in a few weeks,” Parr noted.

Major upgrades were rolled out in December 2015, including an application programming interface that allows users to build their own tools to access USPTO’s assignment system and the data behind it. Users can also now view the underlying legal documents related to their searches, rather than just a coversheet.

The initial project kicked off in 2014 between August and September. As a major transparency initiative, it drew interest from the White House, which gave impetus for a quick completion.

“So we really did fast-track it to get it out by December,” Parr said.

That timeline is rare in government, but there were a number of factors key to the project’s success. White House officials and policy-makers were involved early on, and they supported the project. Parr and his team used agile development to add user capabilities in rapid increments; they worked early and often with people who file assignments and other users to incorporate their feedback; and they were able to use an existing contract rather than creating a new one.

“That was a big help to us,” Parr said. “There was no way we were going to make our Jan. 1 deadline without it.”

6.7
MILLION
RECORDS

6k
USERS
PER WEEK

3k
SEARCHES
PER DAY

4
MONTH
TIMELINE

SUCCESS FACTORS

1
Get high-level officials involved early on.

2
Let the agency’s business units share in the decision-making.

3
Explore agile development & actively participate in daily scrum & project meetings.

Stop Cyber Threats In Their Tracks

Put Big Data Analytics to Work in Your Cyber Environment.

ViON's DataAdapt S2000 Cyber Secure analyzes large volumes of sensor data in near-real time to:

- Identify and thwart cyber attacks
- Eliminate insider threat vulnerabilities
- Provide actionable predictions for Advanced Persistent Threats (APTs)

LEARN MORE

[www.vion.com/Solutions/
Cyber-Security.aspx](http://www.vion.com/Solutions/Cyber-Security.aspx)



How Data Analytics Drives Real-Time and Forensic Analyses

An interview with Billy Hill

Customer Solution Architect for Big Data Analytics at ViON

The data analytics landscape is rapidly evolving thanks to more powerful and affordable tools for data gathering and processing.

Memory, CPUs, and fast storage are getting cheaper as the manufacturing processes are improved. Each generation is becoming more familiar with the ecosystem of analytics and data storage technology, said Billy Hill, Customer Solution Architect for Big Data Analytics at ViON. This combination of affordability and usability bring analytics tools within reach for government agencies, businesses and even consumers.

“Although the technology has been around for years, it’s now accessible to a wider audience,” Hill said.

People are using more CPUs and memory in concert, and these technologies are starting to be incorporated into the workflow of researchers and mathematicians, who ultimately develop the algorithms used to find patterns in the data. Advances in hardware are helping to drive changes in the types of algorithms that are developed. At the same time, the algorithms are also dictating how the hardware is used. “It’s a synergistic development,” Hill said.

As the technology advances and provides better insights, the demand for data analytics and the outcomes they produce continues to increase. “Leaders nowadays are expecting to get more value out of their data,” Hill said.

That’s why agencies are using different types of analytics, such as predictive and prescriptive to better understand what may happen in the future and how they should respond. Depending on the type of analysis they want to conduct — real-time or historical — there are different data gathering and analytics

tools to support those efforts. Let’s take a closer look at each of these examples to better understand what they entail and how they benefit government agencies.

Predictive and prescriptive analytics. It’s hard to say with certainty what will happen a month from now or even a year from now. But agencies can use the data they have to better predict what may happen in the future. That’s the real value of predictive analytics. It helps agencies plan accordingly in a variety of areas, whether it’s budgeting, future customer demands for services or internal human resource planning. But knowing what may happen does an agency no good if it’s unprepared to respond. That’s where prescriptive analytics comes into play.

“If agencies are able to predict the probability of an event or action occurring, that information can be used to drive their decisions,” Hill said. “Prescriptive analytics helps decision-makers figure out what to do with the results from their models. To make decisions you have to be informed, and the better your information, the better your decisions.”

Streaming vs. batch analytics. Streaming and batch refer to the types of analytics technologies that are used to drive specific outcomes. For example, streaming technologies are usually used to power real-time data analysis that require immediate data processing and insights. Streaming technologies involve the use of sensors to rapidly collect the data that agencies must then analyze.

Batch analytics, on the other hand, is used in scenarios that don’t require immediate decision-making. In this instance, agencies have more time to tweak their models for more accurate insights.

“There is a greater focus on quality if you’re doing batch, where as with real-time analysis there are online algorithms making decisions as they’re running,” Hill said.

Real-time vs. forensic analysis. Now that we’ve discussed some of the types of analytics agencies can use, let’s explore the outcomes they produce. To reiterate: Streaming analytics supports real-time analysis, while batch analytics supports forensic analysis.

Both real-time and forensic analyses provide great value to agencies, depending on their specific requirements, Hill said. Real-time analysis is helpful when decisions are based on quickly evolving data, such as current traffic on the road or changing weather conditions. But batch analytics would be the more appropriate option for driving forensic analyses looking at historical data, such as traffic patterns over time or weather conditions over the past week or several months.

In an emergency response situation, first responders would rely on real-time analysis to determine how far an ambulance is from a victim, and how far the victim is from the nearest hospital. A forensic analysis of that same data may be used to better understand which parts of the city receive the most 9-1-1 calls and where ambulances should be stationed to better respond to those calls.

“We’re always interested in things that are just out of reach, and some people think this technology is out of reach — but it’s not,” Hill said. Thanks to data analytics, agencies have all of these capabilities and more at their fingertips.



Government Accountability Office



An interview with Seto Bagdoyan

Director of Audit Services for GAO's Forensic Audits & Investigative Service

THE CHALLENGE: Tracking Medicare Fraud

Tracking down fraud in the government's multibillion-dollar Medicare program is not for the faint of heart.

Consider this: Medicare shelled out \$554 billion for healthcare and other related services in fiscal 2014. But \$60 billion — or about 11 percent — of that amount was paid improperly, according to estimates from the Centers for Medicare & Medicaid Services (CMS).

Ineligible providers and suppliers intent on defrauding the government have used all sorts of schemes to enroll in the Medicare program. Despite CMS' multilayered screening procedures to weed out fraud, some applicants try to exploit a specific weakness in the agency's verification process by listing questionable addresses that are not "operational" to support Medicare-covered items and services, as required by federal regulations.

A [June 2015 report](#) by the Government Accountability Office found addresses listed as legitimate practice locations that were later traced to mailbox rental stores, a construction site, a fast-food franchise and a well-lit but empty office.

"We found that there were some significant issues that the agency and its contractors were missing," said Seto Bagdoyan, Director of Audit Services for GAO's Forensic Audits and Investigative Service, and lead author of the report.

But how do these blatant cases of fraud slip under the radar? According to GAO, one reason is that the software CMS uses to verify addresses does not flag illegitimate locations.

For GAO, reviewing the data and weeding out fraudulent providers and suppliers, particularly those with ineligible addresses, requires months of intense work.

"It's a very methodical, often tedious process to obtain the data, determine its reliability, figure out what sort of matching or mining you're going to perform, and then write the appropriate program using tools such as SAS and Idea to actually perform the analysis, get the initial results, and verify those results through due diligence before reaching your analytical conclusions," Bagdoyan said.

THE SOLUTION

“We’re a very small agency, and my team...we punch way above our weight because we have the determination to do a good job with the data and the tools we have.”

As Bagdoyan mentioned, there’s a lot of preparation behind the scenes before data can be properly analyzed.

His team worked with specialists in GAO’s Applied Research and Method’s group to write code that enabled targeted analysis of questionable provider addresses submitted to Medicare.

The team also used a commercially available software package from the U.S. Postal Service that standardizes addresses and flags locations that aren’t legitimate. Bagdoyan explained that validating these locations is a very deliberate process.

“It’s not like you throw a million or so Medicare provider names into the address database and it all comes out with results,” he said. “You have to code and program the software to accept the database and then bounce them off of each other to throw off the leads that you will then take out, and confirm through what we call tracing and verifying, which is a very tedious exercise. It’s part of due diligence to make sure that the emerging results are indeed results, rather than programming glitches.”

In its Medicare report, GAO parsed nearly 1 million addresses using the USPS database and identified more than 100,000 of them as

potentially ineligible. Using a random sample of flagged addresses, the agency estimated that 23,400 didn’t meet federal requirements.

“That was our deep dive,” Bagdoyan said. GAO also used Google Maps to track down various locations and sent teams from its field offices to visit questionable sites.

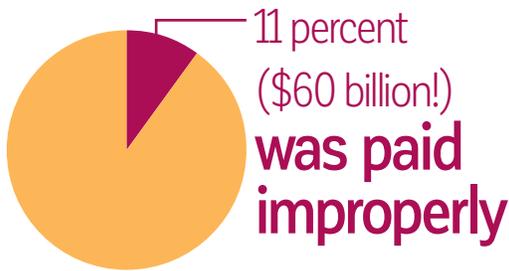
“By looking at a lot of data and drilling down to the appropriate level, we have reasonable assurance that we’ve exhausted what we can do, and we can base any conclusions on the work that we’ve done,” he noted.

GAO successfully combined the power of data analysis software, readily available Internet-based tools and on-the-ground follow-up to verify its findings. It didn’t take an army of experts. In fact, an intern conducted most of the Google Maps analytics work.

“We’re a very small agency, and my team, the Forensic Audits and Investigative Service, is even smaller,” Bagdoyan said. “It’s the smallest mission team in GAO. We’re right around 60 or so people, and we punch way above our weight because we have the determination to do a good job with the data and the tools we have.”

\$ 5 5 4 , 0 0 0 , 0 0 0 , 0 0 0

paid by Medicare for healthcare & related services in FY2014



1 team of GAO employees

sorted through 1 million addresses

and found

23,400

addresses that did not meet federal requirements



SUCCESS FACTORS

1
Ensure data is clean, usable, & reliable as it’s collected.

2
Institute multiple levels of review to double & triple check that the data is secure.

3
Adhere to policies & regulations that govern data collection & publication.



One Flash Platform. **ANY** Workload.

Faster analytics lead to greater insights. Greater insights lead to better decisions. Tegile gives you a multi-tier flash platform capable of supporting high-performance flash and high-density flash to exceed all of your agency's performance and capacity demands.

www.tegile.com/federal

Accelerating Insights with Multi-Tiered Flash Storage

An interview with Chris Tsilipounidakis & Jason Strawderman

Manager of Product Marketing and Vice President, Federal, at Tegile

At its core, big data analytics is about converting data into knowledge that can be used to make informed decisions, which can have a variety of positive implications to government operations.

For instance, big data is being used in cyber defense to accelerate recognition of patterns that represent threats to the network. These insights are enabling organizations of all types to predict and prevent potential cyberattacks before they occur, said Jason Strawderman, Vice President of Federal at Tegile, a California-based company that specializes in next-generation multi-tiered flash storage systems.

“Government healthcare organizations are using big data to manage, analyze, visualize, and extract information that is helping prevent and fight disease,” Strawderman said. “While the benefits of big data in our government are tremendous, the implementation is proving very challenging as legacy infrastructure cannot handle the demanding workload, and disparate systems prevent the sharing of information. Big data is forcing our government to modernize their IT infrastructure and transform culture, policies and processes. These challenges are not just limited to the government; many organizations in the private sector face these issues as well,” Strawderman said.

With this transformation comes the need for a different way of storing and managing big data. Agencies must break down legacy storage infrastructure silos while improving performance and increasing capacity, said Chris Tsilipounidakis, Manager of Product Marketing at Tegile.

Customers are moving away from the idea of purpose-built storage, where they use different vendors to store different types of structured and unstructured data, Tsilipounidakis explained. In the past, organizations typically used certain vendors to store mission-

critical data on more expensive, high-powered storage systems, other vendors to store less important data on lower-cost storage, while using more general-purpose systems for workloads that require an even blend of performance, capacity and economics.

“Big data analytics is causing a massive hyper-consolidation in the datacenter,” he explained. “Hyper-consolidation means that if I’ve got less money, less resources, but more requirements to do the same thing that I was doing before, I need to do it in a consolidated fashion.”

Traditionally, legacy storage systems populated with a mix of solid-state and hard-disk drives have been the go-to for agencies’ storage needs. But to help agencies better meet the workload demands of big data analytics, Tegile is working with them to adopt multi-tiered flash storage.

“With a multi-tiered flash-storage architecture, I can tell CXO’s, Program Managers, and IT Managers that I can give them far better performance than they are getting from their legacy storage systems and at a fraction of the cost. In fact, our new High Density Flash arrays offer flash performance at the same cost per gigabyte of traditional hard drives,” Tsilipounidakis said.

The Tegile multi-tiered flash architecture uses different layers of flash to meet agencies’ varying needs, whether it’s managing structured or unstructured data. This approach provides a flexible, extensible, scalable, affordable, and energy-efficient architecture that can accommodate change rapidly. With dense flash, agencies can get better performance and more data into a smaller footprint.

“Our new high-density flash array, IntelliFlash HD, can pack up to four petabytes of raw flash in a single 42U rack, at near the cost per gigabyte of traditional hard drives, and incredibly power efficient,” said Strawderman.

In terms of data management, security is also top of mind for agencies. That’s why Tegile uses self-encrypting flash drives that are FIPS 140-2 validated, meaning they meet U.S. federal standards for encrypting data at rest and in transit.

“Tegile can adapt to fit any storage environment and workload,” Strawderman said. “Most government agencies are migrating to — or already operating — hybrid cloud architectures consisting of both private cloud and traditional IT. We can interoperate in those existing environments, without them having to rip and replace all the legacy technology.”

When it comes to choosing cloud or traditional IT, it isn’t an all-or-nothing decision, Strawderman added.

However, one area where a model like cloud can be less viable is when agencies want to process data in real-time and take quick actions, as opposed to reviewing large amounts of historical data for less time-sensitive analyses, Tsilipounidakis noted. That’s why agencies must first evaluate their needs and see where different technologies can support those objectives.

Strawderman is an advocate for the so-called “land and expand approach,” where agencies can start by using flash storage for their big data needs and over time expand flash to other workloads, such as virtualization, virtual desktop infrastructure and archives and backup, as older legacy systems are retired and refreshed.

“Nobody wants to replace their entire infrastructure, and no one has the budget for that,” Strawderman said. “But big data analytics, server consolidations, desktop virtualization and mobility projects are a few great opportunities to introduce flash into their environment.”

General Services Administration



An interview with Paul Tsagaroulis

Director of Human Capital Analytics at GSA

THE CHALLENGE: Improving Human Capital Management

Data can tell us a lot about the government workforce — who was hired, who left, why they may have left, and who else is likely to leave.

But it's hard to glean those insights with data drawn from disparate systems, in different formats. And those who need the data the most can't easily access it, said Paul Tsagaroulis, Director of Human Capital Analytics at GSA.

Although GSA has a single human resources system, it comprises multiple subsystems that don't talk to each other. This only exacerbates the challenge of combining disparate data sets to tell a complete story. To stay current, GSA has had to build separate processes to capture new HR data sets, such as telework data.

"A big challenge for us is data management and information management," Tsagaroulis explained. "We spend a lot of our time focusing on getting access to the data, making sure it's valid and reliable, and that leaves very little time for us to analyze the data and look at trends and patterns, maybe do some sophisticated analytics, and some predictive and prescriptive analyses."

But that's changing, thanks to the work Tsagaroulis and his team oversees.

Prior to GSA's 2013 [consolidation of its HR functions](#), inefficiency was a big problem. There were offices within each of GSA's business lines, and they all performed the same data analytics services, multiple times, for different agency customers.

"GSA consolidated, and now we've been able to take a top-down approach in this process," Tsagaroulis said. "So we do it once, we deliver it once, but we can adjust it for any business line's specific needs, geographic needs or a combination of those. We were created as a division to support this for the agency, to alleviate the need for business lines to do this themselves."

THE SOLUTION

“Looking further into turnover data, we can predict with some confidence the number of people we would expect to have retire or voluntarily separate from the agency.”

In 2014, GSA launched a division dedicated to developing workforce analytics deliverables. The first of those deliverables was an interactive dashboard that visualized answers to common HR questions about new hires and separations by grade, onboarding by major job series, and more.

“We dramatically reduced ad hoc reporting and responding to requests as they came in,” Tsagaroulis said.

The dashboard data has already influenced decision-making at the administrator level. After cutting the number of annual external hires from about 800 in fiscal years 2011 and 2012 to about 400 in 2013, GSA was in a bind.

The agency filled fewer positions but at higher General Schedule grade levels, which meant higher costs. The dashboard confirmed what many suspected: By hiring people in more senior grades, the agency was increasing its costs and diminishing its talent development pipeline of younger hires coming in at lower grades. To address the imbalance, GSA implemented an entry-level hiring program. In fiscal 2015, the agency tripled the number of employees it hired at the GS-7 level, compared with the previous year.

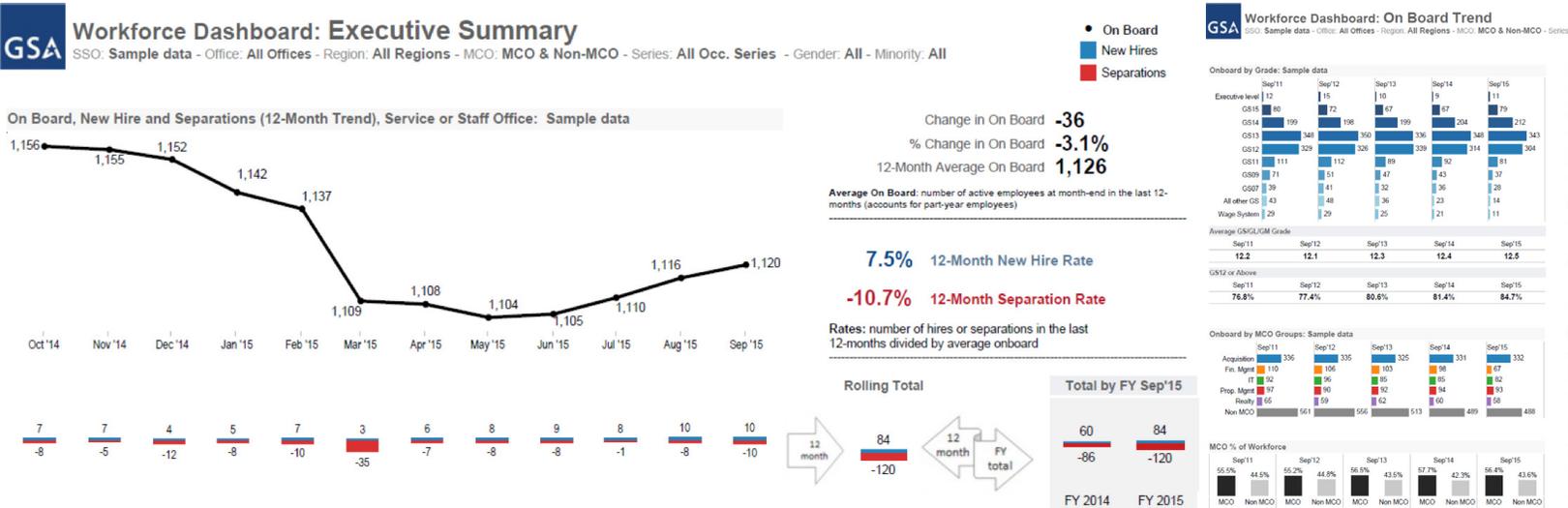
“The dashboard was useful in starting the conversation,” Tsagaroulis said.

Early on, his team examined trends across GSA's workforce data. Reviewing those trends led the team to ask specific questions about turnover rates, what factors correlate to turnover, and if GSA could predict turnover. This led in turn to the development of a turnover model for predicting future employee losses.

“One of the findings that we came across was that people in the central office in the D.C. area are more likely to leave for another agency, whereas individuals in the regions are more likely to retire,” Tsagaroulis explained. “So we can account for some differences in turnover based on geographic variables. Then, looking further into turnover data, we can predict with some confidence the number of people we would expect to have retire or voluntarily separate from the agency, and we could calculate the cost of the attrition savings over the course of the year.”

From there, the agency can use those estimates to do some recruiting ahead of time, in anticipation of turnover.

Overall, the division has improved how GSA manages its HR data sets by automating reporting of turnover, onboarding, and performance data. Building that foundation has made it easier for the division to employ prescriptive analytics and help managers develop action plans for staffing. Tsagaroulis' team is in the process of rolling out an internal HR data warehouse to consolidate data and improve data sharing.



SUCCESS FACTORS

1 Tackle data management issues first.

2 Use visualization techniques to explain trends & other findings in the data.

3 Let workforce data dictate new HR solutions or programs that maximize employee potential.



Beyond Big Data to Data-Driven Insights

More effective government sits at the intersection of insight and action.

When you can collect, unify, and analyze all the data that surrounds your department, government is empowered to uncover key insights that matter most. It could be an insight that changes the way citizens view your agency, or how your department sees their role. Learn how big data and analytics for smart government can help you know more so you can do more in this informative **white paper**.

Whitepaper: Analytics in Service of Smart Government

What would you do if you knew?™

TERADATA

The Future of Data Analytics in Government IT

An interview with Alan Ford

Director of Presales Consulting, Teradata Government Systems

What does 2016 look like for government IT? We may be in for a few surprises, but one thing is clear: data analytics will play a key role. There are several areas where data and analytics will drive more value in government, including data transparency, tax fraud, and transportation logistics.

Data analytics will require a continued effort in analyzing increasingly larger datasets. And these larger datasets are coming at government agencies at greater speeds than ever before, with wider variety and complexity.

This is where challenges begin to occur. With more data to analyze, government agencies can easily be overwhelmed as they try to maintain system performance and data integrity while protecting sensitive information. Additionally, with most agencies having legacy systems in place, it can be difficult to integrate new systems and different vendors into their existing analytic ecosystem.

There are ways to combat these challenges. In an interview with GovLoop, Alan Ford, Director of Presales Consulting for Teradata Government Systems, shared how agencies can leverage data analytics, including those in the cloud, to generate actionable information and enable better data sharing within and across agencies.

Teradata provides end-to-end solutions and services like data warehousing to enable government agencies to create unified data ecosystems. The Teradata Unified Data Architecture, for example, integrates different platforms into a comprehensive analytics solution that enables fast, deep, and powerful data management, storage, and exploration. With this architecture and its inherent analytic capability, agencies can complement and augment their legacy systems rather than abandoning them.

These technologies will help all areas of government. Ford cited healthcare fraud as a major data analytics problem for government. Healthcare comprises some of the largest annual expenditures in both the private and public sector. Medicare and Medicaid, in particular, receive a significant amount of government funding, which makes it an attractive target for fraudulent activity.

According to Ford, however, there is hope with increased data digitization and analytics automation. "Digitization has created a new way of automating fraud detection," Ford said. "By combining traditional claims data with other separate but publicly available data – like bankruptcies, liens, judgments, or even repeated address changes – into an integrated data model or a unified data architecture, we can run advanced analytics that make fraud easier to track or even proactively predict the likelihood of its occurrence."

Teradata's Aster Discovery Platform is one tool that helps run analytics that could only be done manually in the past. "With this tool, we can run older transcribed handwritten records or typed reports through complex text analytics capabilities and sentiment analysis in an automated fashion dramatically decreasing the time to results," Ford said. "And with less expensive storage accessible by sophisticated techniques, all within the unified data architecture, we'll be able to analyze all the data rather than just representative samples of it."

When looking to the future, Ford predicts data analytics and cloud will make a strong pairing for government. Governments will benefit from the reduced costs enabled by such technologies, so agencies can do more with less.

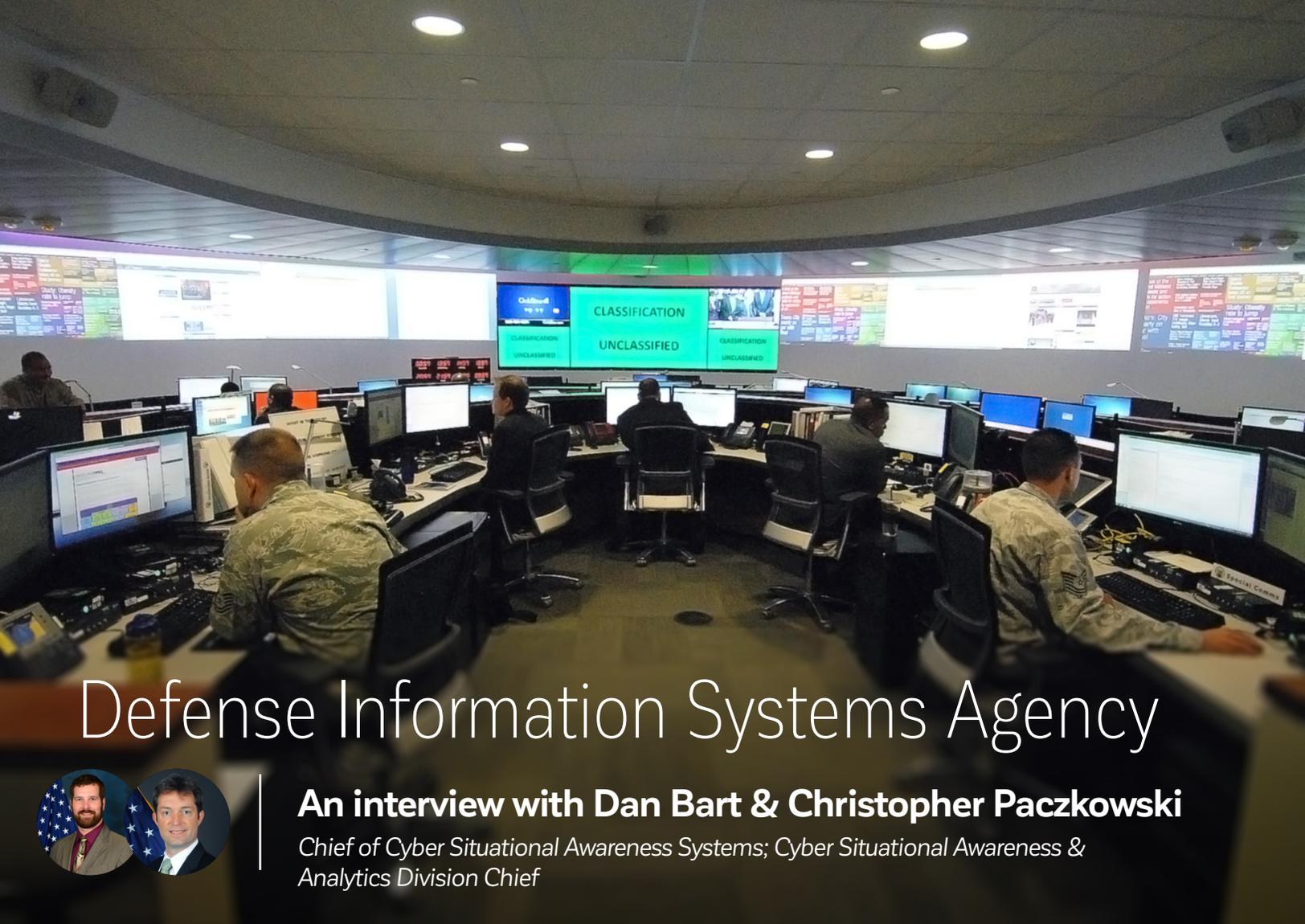
"The whole movement into the cloud environment is a huge cost saver, because no longer do organizations have to set up and maintain their own infrastructure," Ford said. "It's taken care of by the cloud provider. In this way, agencies convert acquisition costs into operating costs."

This big leap to cloud and data analytics, however, may seem improbable for government agencies with legacy systems. When agencies have spent a tremendous amount of money acquiring systems over time, they are reluctant to replace these familiar platforms.

That's where Ford says "the Holy Grail of analytics" – the unified data architecture – can come in handy. "This is an ecosystem which allows for the harnessing of all of an agency's data," he said. "We also refer to it as the logical data warehouse because it brings the right type of analytical processing to the underlying datasets. Further, it allows users to combine and analyze data from multiple disparate platforms all in a single query."

A logical data warehouse can subsume older platforms so agencies that choose not to transform and migrate data into an integrated data model may still leverage their legacy systems.

With a unified data architecture that can accompany legacy systems, complemented by cloud technologies, government agencies can better manage the vast amounts of data accelerating toward them. The future of data analytics does not have to be daunting. With a unified data architecture, inside or outside of the cloud, government can be ready for advanced analytical capabilities and harness large influxes of complex data to further agency initiatives.



Defense Information Systems Agency



An interview with Dan Bart & Christopher Paczkowski

Chief of Cyber Situational Awareness Systems; Cyber Situational Awareness & Analytics Division Chief

THE CHALLENGE: Empowering Cybersecurity Analysts

Imagine that you're an analyst at the Defense Information Systems Agency. Your job is to detect and investigate anomalous network traffic that may be nefarious, including malicious insiders and nation states seeking to steal classified data and wreak havoc on defense systems.

Keep in mind that in fiscal 2014 alone, federal agencies reported nearly 70,000 security incidents. But that number doesn't include the billions of unsuccessful hack attempts that agencies had to thwart each month.

As if an analyst's work weren't hard enough, imagine having to monitor myriad systems, looking for red flags that may or may not be the work of bad actors. Not that long ago, DISA analysts used a manual process to see the anomalous behavior and connections between various system alerts.

"In a lot of cases, it's either missed or they get tips and cues from something else," alerting them to dig deeper, said Dan Bart, Chief of Cyber Situational Awareness Systems. "For example, if a firewall log alerted an analyst to anomalous activities, he or she could then open

an investigation and look at other sensors to try to draw connections from the data."

That was the old way of doing business.

About 18 months ago, DISA deployed a new anomaly detection system that in some respects will provide those connections for analysts. Rather than sending alerts that something may not be right, this new system will let analysts know when something is not right, Bart said.

Consider how many end users likely log on to their computers outside of normal business hours. With the rise of telework and flexible schedules, it's hard to say what a "normal" workday looks like for any given employee.

But analysts have to distinguish between anomalous activity that is normal and abnormal.

THE SOLUTION

“If you have a number of different log alerts or data sets that are flagged, then you’re able to correlate those and start answering heuristic questions.”

What’s far more valuable for analysts is the ability to easily correlate multiple data sets, such as what time users are logging on to their computers, how that compares with their normal computer use, and what they were downloading or uploading, said Paczkowski.

“You can start to add A, B, and C together to say, ‘Alright, this indicator is a lot higher priority than the rest of them,’” Paczkowski added. “So now you’re giving the analyst an ordered list of priorities to look into,” instead of following up on every user who logged in after regular business hours.

Analysts can work more efficiently because they aren’t flooded with a sea of alerts that realistically can’t all be addressed at the same time. This system has helped DISA prioritize anomalies and understand what data is leaving the network in an unauthorized way, Bart said.

The genesis of the program has some ties to the devastating leaks by former government contractor Edward Snowden about the National Security Agency’s controversial data collection processes.

But it isn’t just insiders that the system can detect. The same anomaly detection can be used to identify a nation state worming its way into the network, or even a machine misconfiguration.

“So if you have a number of different log alerts or data sets that are

flagged, then you’re able to correlate those and start answering heuristic questions,” or noting patterns in behavior, Bart said.

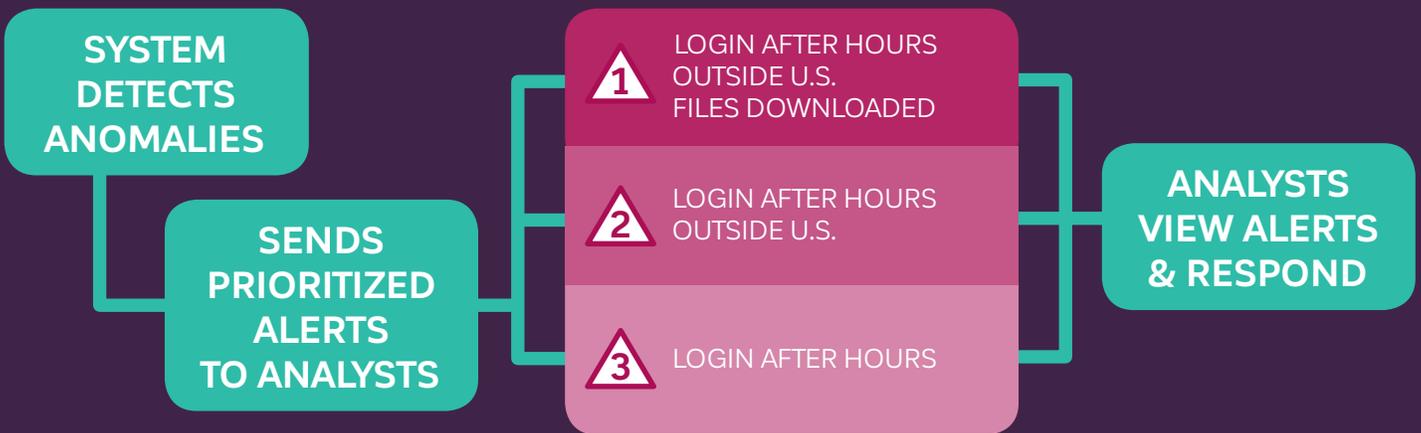
For example, if an analyst sees the same pattern with a user logging in at an odd time or from an odd place, the system provides an analytic tip or alert that will ultimately lead to a countermeasure. Humans decide what follow-up measures to take — not the system.

“It doesn’t solve it,” Bart said. “It doesn’t catch the guy, but it does tip and cue to say, ‘Hey, this should be looked at.’”

Since its rollout about 18 months ago, the system continues to evolve. The agency is working to fine-tune the algorithm that defines and prioritizes threat factors. Think back to the earlier example of the individual who repeatedly logs in after hours, at different times. While it may raise eyebrows, it’s not uncommon or particularly suspicious.

If analysts combine that algorithm with different priorities such as data exfiltration, where the person logged in, what devices they used to connect to the network, and other behavior, then they can combine that information to get a prioritized rank of anomalous behavior.

“Alerts used to be more compartmentalized, based on specific behaviors,” Paczkowski said. But now the focus is on tying that picture together to give analysts a more complete threat profile.



SUCCESS FACTORS

1
Let the use case dictate what sensors are needed and what data should be collected.

2
Communicate with other people in the department, and get a feel for what systems and capabilities they’re using.

3
Develop analytics in a common way, and work toward a common goal across the department, to create opportunities to share analytics.

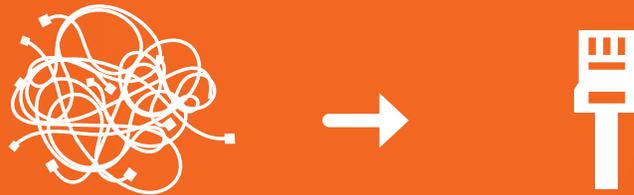
When government runs on flash



What took days, now takes hours



What filled data center racks,
now fits in a box



What was complex, now is simple



What became outdated, now is evergreen



Pure Storage (PSTG) is the #1 recommended technology vendor (based on Satmetrix 2015 NPS Global B2B Benchmarks). To learn more about how Pure Storage can help your agency accelerate what's possible, visit www.purestorage.com/government

© 2015 Pure Storage, Inc. All rights reserved. Pure Storage, Pure1, and the P Logo are trademarks of Pure Storage, Inc. All other trademarks are the property of their respective owners.

In the World of Data Analytics, Speed Is Everything

An interview with Vaughn Stewart

Vice President, Enterprise Architect, Pure Storage

The speed at which government agencies can access, analyze and disseminate information impacts more than cost savings and internal efficiencies.

For officers in Garland, Texas, it's about the ability to quickly run queries on individuals during routine traffic stops and determine what actions to take based on that information. In many states, government officials must be able to quickly detect and assess deteriorating weather conditions and take necessary precautions, such as closing roads or temporarily reducing driving speeds.

But how can agencies support the growing demands for faster services when budgets are limited? For a growing number of agencies, the answer to powering these types of operations has come in the form of flash storage, which can now be a more cost-effective and efficient alternative to the legacy disk storage traditionally used by agencies.

"The power of flash really allows agencies to advance their ability to gather information," said Vaughn Stewart, VP, Enterprise Architect at Pure Storage. "It's focused around collecting information faster to they can respond more quickly to different scenarios, such as emergency response, cybersecurity or in the financial sector to detect and prevent fraud."

The ability to process more data faster allows agencies to run more queries, gain greater insight, and have a deeper knowledge of their growing data, Stewart explained.

Much of that data, however, is stored in various formats across different systems, which can put a strain on data sharing.

Although there are many technologies that can provide a central repository for data, the speed at which agencies can process that data is another factor agencies must consider. That's where flash storage has come in as the game changer for agencies that are open to change.

"Many agencies don't look at flash until they have a known issue, and they've exhausted all other technologies," Stewart said. "Usually that's because flash has historically been cost prohibitive. But the reality is that we are in an age where flash is not just affordable, but in many circumstances, it costs less than spinning disk storage."

With this in mind, Stewart said some agencies are starting to consider flash earlier on, rather than as a last resort to resolve their biggest pain points. Forward-thinking customers understand that moving to flash can help them speed their work because the latency issues often associated with disk storage won't plague them.

"For every month that you run on flash, you can complete a year's worth of analytics that are completed on disk," Stewart said.

There's one case in particular Stewart noted that involved data scientists looking for fraud. At first, they could only operate four ad hoc queries at a time using disk arrays. But when they moved to flash storage, that number jumped from four to 200 queries.

"Data storage is like a foundational layer underneath all of these systems," Stewart said. "Think about a building. If you've got a poor foundation, you can only build a building

to a certain height, size and scale. If you can replace a building's foundation with a new generation of technology or architecture technology, you can then allow that building to scale higher. Data storage is very similar."

If agencies switch the infrastructure from spinning disk to flash media or solid state storage, it allows them do exponentially more with that same set of IT services running on top of the storage, Stewart added.

"At Pure Storage, we can provide a solution that can handle 10 times the capacity at about the same cost of overhead as a spinning disk solution," Stewart said. "That allows agencies to keep up with the demand without having to come up with additional funding. With a business model called Evergreen Storage, Pure offers flat annual maintenance and allows agencies to stay modern by upgrading software and controllers (free of charge) every three years." This model gives agencies tremendous flexibility and peace of mind.

"Analytics brings to the table an improved end user experience because it's a proactive model for what is happening, trending or changing, and how to immediately implement a change or action," Stewart said.

With Pure Storage, agencies not only get the benefit of flash storage but also seamless performance that is not impacted by maintenance or even upgrades. It's a much simpler (and less expensive) alternative to the industry norm, where agencies have to rip and replace storage regularly due to spectacular increases in maintenance costs.



Norcross Police Department



An interview with Bill Grogan

Support Services Captain at Norcross Police Department in Georgia

THE CHALLENGE: Predicting & Reducing Crime

No police department has the manpower to be everywhere at all times. Even in a city the size of Norcross, Ga., which spans only 6.5 square miles, that's still impossible.

Some 16,000 people call Norcross home, but nobody knows how many come into the city every day for work. "The problem that we have is, for such a small area, we have quite a number of armed robberies and burglaries," said Bill Grogan, Captain of Support Services for the Norcross Police Department. "I don't have the resources to cover 6.5 square miles, 24 hours a day, seven days a week."

Like many other law enforcement agencies, the department has relied on heat maps to pinpoint hot spots for crime. "It's old school, in my opinion," Grogan said. "It's too convoluted for the average officer to understand where he or she needs to be."

Plus, these maps can only give officers a historical view of where crime has occurred. That kind of data is fine for viewing trends over time,

but for officers patrolling the streets, the solution is reactive — not proactive.

"We wanted to find something to lessen the number of property crimes," Grogan said. But the department wasn't searching for just any solution. It wanted to invest in resources that could proactively show officers where crime is most probable.

In 2012, a new police chief took the reins and shared his vision of using technology to make everyone's job easier, not more cumbersome. About a year later, both Grogan and Police Chief Warren Summers stumbled upon an article about the use of predictive analytics in law enforcement.

"We wondered if it was possible to predict crime," Grogan said.

THE SOLUTION

“This is not a crystal ball or ‘Minority Report.’”

That question was put to rest on the first day the department rolled out predictive policing software in 2013. Officers nabbed someone wanted for burglary, and a few weeks later they caught burglars breaking into a home. Those are just a few of the success stories.

This is how it works: The company that provides the software gets real-time data feeds and historical data from the department’s records management system, including the type of crime, where it happened, and the date and time it occurred. Using algorithms and very fast data crunching, the software generates “prediction boxes” that show officers the most probable location where a crime might occur.

“This is not a crystal ball or ‘Minority Report,’” said Grogan, referring to the science-fiction movie in which police used technology to arrest murderers before they committed crimes.

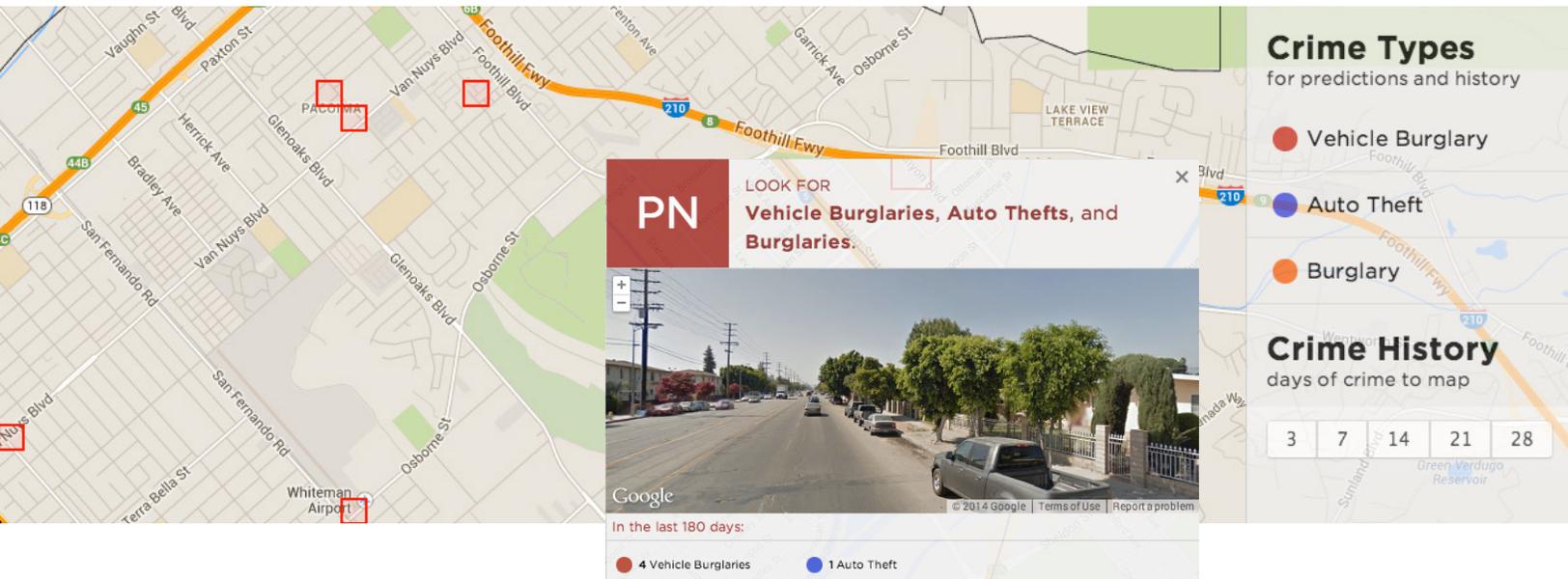
The data that officers receive from the software tool is in real time. When they first come in for their 12-hour shift, officers log on to their computers to view the prediction boxes — eight red boxes, each representing a 500-by-500-foot area, spread over a map. Officers can zoom in on the boxes, which might encompass an apartment building, a gas station or a roadway.

All of the department’s police cars are equipped with GPS, and supervisors can monitor the amount of time that officers spend in the boxes. If officers spend too much time in the boxes, they’re neglecting other parts of the city. The goal is to spend a little more time in the prediction boxes than in other areas because that increases the chance of preventing crime, Grogan said. Those boxes change for every officer, depending on the time of day, the most recent crime data, and the type of crimes the department wants to target.

It took more than a year to get everyone on board with the software, Grogan said. It was easier to get new officers to use the software because they had no other routine coming in to the department. But veteran officers were another story. “Police officers do not like change. We prefer things go like they’ve been going,” he said.

In addition to tracking the extent to which officers monitor their prediction boxes, the department also tracks whether or not officers use the software.

Grogan said officers have become increasingly good at recognizing areas where there’s a higher probability of crime. The technology is easy to use and requires very little training, which helped boost adoption.



SUCCESS FACTORS

1 Properly train end users to use any new analytics software.

2 Seek their feedback on how to use the tool and improve processes.

3 Work with the data analytics provider to share valuable feedback on how to improve the tool.

WHEREVER BUSINESS TAKES YOU, YOUR DATA IS READY. WELCOME TO DATA FABRIC

Manage your data seamlessly across any environment and move it to where you need it most. The future of data management is here.

netapp.com/datafabric

 **NetApp**[®]

4 Keys to a Successful Data Analytics Program

An interview with Gus Horn

Global Consulting System Engineer, NetApp

There's no denying that big data analytics has the power to transform the way government operates and serves its citizens. There are countless stories of agencies using analytics to speed emergency response times, improve online services and better track internal metrics.

But the path to a successful big data analytics implementation is one that requires proper long-term planning, said Gus Horn, Global Consulting System Engineer at NetApp.

"The reason is there are many variables involved in acquiring and ingesting data into big data analytics platforms, and then there's also the logic around the analytic component, how to write it and gain insight into the data," Horn said. "The biggest stumbling block is typically around ingesting the data into an analytic platform."

It's common for organizations to start out with small development and testing environments and a handful of servers to demonstrate a proof of concept. Although these small environments are a great starting point for agencies to demonstrate the power of big data analytics, the environment must be adapted to properly plan for future data growth and larger projects.

"What we find is that these very small environments are extremely powerful," Horn said. "So it can be very alluring for our customers worldwide, regardless of sector, to start down that path and say, 'This works well, I'm just going to repeat and do more of the same.' They become trapped in an architecture that is extremely rigid in its design, and two or three years down the road they're unable to evolve, respond to newer technologies, or implement new methodologies in these platforms."

The problems typically don't come until later on when the system has grown exponentially, servers age and must be replaced, and disk failures are a constant issue. By that time, it's very difficult to move the data to a new hardware infrastructure, and the employees who originally built the system have long left the organization. At that point, keeping the system up and running becomes very expensive.

To avoid these issues, Horn recommended that agencies plan accordingly in these key areas:

Growth. "Organizations always have to plan for growth when they start talking about a big data platform," Horn said. "Big data platforms by their nature are designed to accommodate the velocity, variety and variability of content, and because of those three V's these systems are never static. So while agencies start out small, they should always be cognizant of the fact that technology evolves."

The central processing units that agencies use are rapidly evolving and so are the storage technologies. But the challenge is these two technologies that play integral roles in supporting big data systems are on two different growth paths. Horn suggested that agencies first understand the problem they're trying to address, as well as their growth demands each month, each year and even further out.

"NetApp's architecture is based on directly attaching the storage to the compute, but it gives agencies the flexibility to decouple it and change out the computational element without having to migrate or rebalance that storage," Horn said.

Flexibility. Agencies that don't consider flexibility early on may find themselves locked into

what Horn calls "rigid architectural designs." In this scenario, agencies will find it harder to replace their servers when needed, without having to migrate content. With NetApp, agencies can keep their data put when updating their technology clusters.

Agility. Not only are flexible compute and storage important, but so is agility in providing different kinds of storage subsystems. One example is the ability to move different types of data to the most economical storage options, depending on how often the data is accessed and its importance. "There's no point in keeping data that's 10 years old and not very relevant in your hot bucket, when you can move that into a cold or glacier-like storage device and still have access to it should you ever need it. You can reserve the high-performance flash technologies for more meaningful data," he said.

One of the perks NetApp provides its customers is the ability to automate data migration from higher-priced storage to lower-priced storage. NetApp works with agencies to tag the data so they can track its usage and determine which storage option is most appropriate.

Evolution. "Oftentimes, agencies get stuck repeating the same mistakes because the task of moving to a newer technology is perceived to be too heavy of a lift," Horn said. Using NetApp's data lifecycle management tool, the company can easily integrate into an agency's existing technology cluster, migrate the data to new hardware and decommission older technology gracefully.

To learn more, visit: <http://www.netapp.com/us/solutions/big-data/index.aspx>



U.S. Postal Service



An interview with Dan Houston Jr.

Manager of Data Science and Exploration, U.S. Postal Service

THE CHALLENGE: Improving Mail Delivery

Few of us stop to consider the massive data analytics operations that happen behind the scenes every time we send and receive mail.

In a single day, the U.S. Postal Service processes and delivers more than 500 million mailpieces. That's more than 350,000 mailpieces processed each minute, on average, or nearly 6,000 each second, and it's only a small slice of the data USPS collects and analyzes. The agency relies on multiple data sets and tools to analyze operational efficiencies and better predict what resources are needed to meet future demands.

For USPS, the problem isn't a lack of data. In fact, USPS data is growing by about 30 percent each year, said Dan Houston Jr., who serves as the agency's Manager of Data Science and Exploration. In that role, he oversees 35 petabytes of data — up from 22 petabytes in 2012 — along with 3,300 databases and 400 applications.

The real issue is being able to quickly pull together large data sets, analyze them and then answer business questions.

"I started two years ago looking at gaps we had from a big data and analytics perspective," Houston said. "We have a lot of traditional technologies, a lot of traditional relational databases, and those work well to a certain point. But when you get to the larger data sets and want to be able to build models across years of data, they don't work very well."

Data provisioning was a major pain point. For example, if someone wanted to know the best marketing campaign for a specific target audience, or information about workloads in USPS mail processing plants, "it always turned into an eight- or nine-week data provisioning exercise to be able to work with that data," Houston said.

The hard part was gathering the data in one place so people could work collectively to provide data-driven answers quickly, rather than nine weeks later.

By then, "we might have missed the opportunity," Houston said.

THE SOLUTION

“By indexing and ingesting all of that data in one place, we can quickly see patterns too. Not only do you have the data in one place but you can also quickly get visualizations, and those are just out of the box.”

To address its data provisioning problems, USPS turned to the open source software Hadoop. Houston said the agency was looking for a platform that could store all types of data, and later format that data when a query is made — a process known as late binding.

But once the data is together in one place, how do you quickly visualize the data and act on it? That’s where the agency’s indexing and search tool, which integrates with the Hadoop environment, comes into play.

“By indexing and ingesting all of that data in one place, we can quickly see patterns too,” Houston said. “So that’s one of the great powers of all of these tools. Not only do you have the data in one place — and you can certainly look at that raw data — but you can also quickly get visualizations, and those are just out of the box. You don’t need a bunch of developers to start developing dashboards right away. You can get some quick visuals and explore the data quickly before you move on to developing an application.”

USPS uses these visualization tools in various ways. On the cybersecurity front, they help the agency better understand where it’s potentially exposed to attacks. It’s one thing to say USPS thwarted 40,000 attacks, but it’s far more powerful when people can see a map of where those attacks originated and what parts of the environment attackers were trying to enter.

“The other big thing we’re working on right now is predictive workloads, being able to better tell our plant managers and plant workers not just the volume of pieces they’re going to receive, but also what equipment those pieces are going to have to be run through to get through their facility — what pieces are going to go out to local post offices, and what pieces are going to go to another plant for further processing,” Houston said.

Having this level of information will ideally help managers improve employee scheduling. The goal is to provide employees with a 10-day work schedule, rather than changing schedules daily. Using the data, the agency can predict the path a mailpiece or package will take through a plant and better address defects that may have thrown that package off course, such as mail volume, a broken machine or an issue with the sorting equipment.

Dynamic routing is another area of focus for USPS. The goal is to find the optimal route for mail carriers, which may require adjusting routes in real time using GPS and other sensor data. Let’s say one carrier has a big load and falls behind on some deliveries; dynamic routing would allow the agency to shift some of that load to another carrier nearby who has a lighter load.

USPS IN A DAY

512

million
mailpieces
processed
& delivered



356,103
a minute

5,935
a second

\$223.7

million
in revenue

320,132

money orders issued

113,531

address changes processed

17,029

passport applications accepted

SUCCESS FACTORS

1

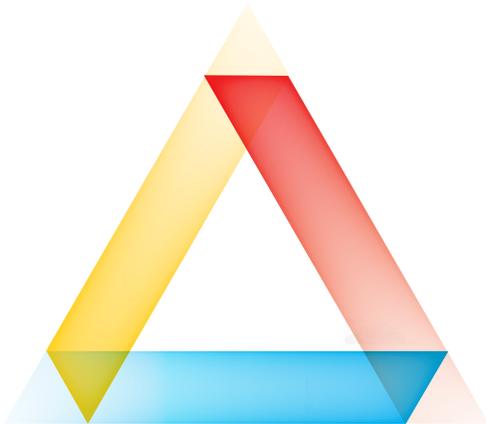
To reduce skepticism, find a pain point that’s relevant to people and show them how data can make it better.

Make data analytics tools and data accessible and usable to a wider audience, not just people with advanced programming experience.

2

Focus on making data operational, so people can quickly visualize the information and act on it.

3



Transforming Data Into Customer Value

-  Customer-Facing Analytics
-  Differentiated User Experience
-  Information as an App



informationbuilders.com

Get Social

WebFOCUS iWay Software Omni



Data Analytics as an Asset

An interview with Dr. Rado Kotorov

Chief Innovation Officer and Vice President, Information Builders

Data analytics has long been a buzzword for government and the public sector. It's used in a number of ways to better serve citizens. But despite its growing popularity, government continues to face many challenges with data analytics. Given the vast quantity of data out there and the cost of management and analytical processes, many agencies have yet to transform data analytics into a real asset.

Why is this? First, government organizations face a huge challenge when it comes to integration and interconnectedness of data sources. Different data are stored in different data sources in a variety of formats, which makes it especially difficult and expensive to access, translate and analyze. The more variety of data sources and formats, the more costly it is for government to run analytics across various organizations.

Second, there's the difficulty of properly publishing and disseminating data to their constituents – whether it's the public, partners or customers. Transforming data to serve in a public, self-service capacity requires new technologies and new, innovative formats. When governments fail to provide accessible data to their constituents, they lose accountability and transparency in the public eye.

So how does government tackle these challenges and use data analytics to turn data into an asset? Dr. Rado Kotorov, Chief Innovation Officer and Vice President of Information Builders, shared how government can navigate these data analytics challenges and how Information Builders can help.

Information Builders is in the business of providing full information lifecycle management, which includes the tools to access, manage, analyze and publish data. "Where we help the most is providing platforms to publish this infor-

mation to citizens and end users," Kotorov said.

These platforms allow consumers to interact with the data, rearrange it, and customize it without having to learn new tools or install anything new on their machines.

"We try to make it easier for organizations to distribute analytical information to the end user without requiring any effort on the user's part," Kotorov said.

Sounds like a quick fix for these challenges, but what makes data analytics more complicated is that solutions often simply can't communicate with each other. Kotorov explained the difficulty of achieving interconnected data through the concept of data warehouses versus data lakes.

"You have information in one data warehouse and different information in another data warehouse that is physically separated," he said. "With a physical data warehouse, you could have many individual data warehouses and data marts and even many more Excel sheets and other personal data files."

These physical siloes make it difficult to have seamless communication between an organization's various levels of data.

That's why many are turning to a data lake, which is one big data storage pool that houses different forms of data. "The data lake is a logical concept, that is why some people call it the logical data warehouse," Kotorov said. "This approach helps you eliminate the silos, but it requires a very flexible integration technology that allows you to quickly access the different sections of the pool of data."

In this manner, the data lake becomes a logical structure that maps how you can connect different sources and how you can get data

from each source and combine them together. That's why Information Builders also helps organizations combine their data warehouses into a data lake.

Creating data lakes, however, is only part of the solution. Kotorov says government also needs to focus on taking the following steps in order to fully utilize data as an asset:

1. Integrate different data sources within data lakes. This enables your organization to get more in-depth and seamless information from your data.

2. Maintain the data. Maintaining your data for data analytics is like maintaining the quality of a house. It needs to be routinely monitored, checked for problems, and repaired accordingly. This ensures longer data shelf life.

3. Analyze and publish your data. Organizations need to ensure they have the analytical capabilities to analyze and publish their data. When striving toward the final product, always keep the end user in mind. Manage the information to turn it into a product consumable by the citizen or end customer.

Stronger data analytics is done best when considering the ultimate consumer. That's why Information Builders works to provide the tools and platforms necessary for government to facilitate better analysis, integration, publication and dissemination of data. This is critical to making data analytics more valuable to government agencies, more accessible to citizens, and ultimately, ensuring more trust between governments and their constituents.

What's most important is that government continues to keep the consumer in mind when transforming data analytics from a formidable challenge into a valuable asset.

Pacific Northwest National Laboratory



An interview with Justin Brown

IT Engineer at Pacific Northwest National Laboratory

THE CHALLENGE: Enhancing IT Investment Tracking

When you're supporting scientific teams focused on preventing acts of terrorism and the non-proliferation of weapons of mass destruction, you have to get it right. Ensuring those researchers have uninterrupted access to functional workstations and online resources is critical.

"Our job is just to do what we can to support and enable that mission," said IT Engineer Justin Brown of the Pacific Northwest National Laboratory (PNNL). "We wanted to go beyond just limiting the impact on customers when we had problems. We wanted to improve the customer experience because the more time that we save for our customers — the researchers — the more it enables them to focus on the lab mission."

The challenge for Brown and his team is that different departments within the larger IT organization collect data for different users, but there wasn't always a consolidated view of that data.

Having the data together in one place has improved operations. Now everyone has access to it and can do correlations across different departments, Brown said. But getting there was a process.

It isn't easy to do that when each group within the IT department has its own tools. For example, the network team has analytics tools for monitoring its devices, the desktop team has tools for monitoring logs and issues with the workstations, and the server administrators have tools as well.

"If one service was affecting another, having that visibility or access to those tools wasn't always there," Brown said.

The CIO put in place a team to focus on improving IT operations through automation and monitoring. Prior to creating the five-person Chief Infrastructure Architecture team, these engineers were bogged down with more repetitive and administrative tasks.

THE SOLUTION

“Once we started seeing the potential of that data for the IT support side of things and the operational side of things, we just started heavily digging into it and seeing what we could do.”

PNNL is now using data analytics to glean valuable feedback about the operation and condition of the lab’s IT sources and how well they’re serving end users. But to understand how much data analytics has enhanced IT operations at PNNL, you have to start with the data.

In the past, the IT performance data existed but it wasn’t accessible departmentwide. The difference today is that when the department monitors the performance and usability of servers, webpages and other IT assets, it can also search for correlations and trends.

Having access to the data makes it easier to perform problem analyses, Brown said. No longer does IT have to log in to each server and comb through IT logs to figure out what happened and where things went wrong. If there’s a network glitch, the department can easily review its data for any anomalies or issues that occurred around the same time.

It’s also easier to see if there’s a spike in account lockouts, or software patches that aren’t installing properly. “We can stop those problems before they get too big,” Brown said.

As the department’s monitoring and automation capabilities improve, customer service remains at the forefront. For instance, rather than emailing the whole lab about a website that’s going to be offline, Brown and his team can use more targeted communications to inform the specific users of the site.

The team can use data to determine peak use times and schedule maintenance around that. But the user experience doesn’t stop there.

The data also shows how many clicks it takes users to find what they’re looking for online. Brown sees this insight shaping user interface improvements in the future.

IT purchasing is another area ripe for data analytics at PNNL. Here’s why: Using data about the age of a system, how it’s performing, how it’s loading, and how often it’s used, the department can assign each of its IT resources a risk score. This allows the team to proactively determine what should be refreshed or where additional resources are needed.

“In the past, probably our biggest notification of a problem was when customers called the help desk,” Brown said. “If we have to wait for that notification, then we’ve failed.”

Customers aren’t the only ones who benefit from the department’s investments.

In fiscal 2015, Brown estimates that this department saved about 1,800 hours in efficiencies and \$453,000 through the use of data analytics and process improvements.

When Brown and his team began rolling out data analytics capabilities about two years ago, the focus was on empowering cyberteams to search through IT logs faster.

“But then, once we started seeing the potential of that data for the IT support side of things and the operational side of things, we just started heavily digging into it and seeing what we could do,” Brown said.



SUCCESS FACTORS

1 Determine key performance indicators, or what metrics will be measured. For each service, determine what matters most to the organization and begin monitoring those issues.

2 Develop a methodology or a systematic approach to using the data. Without that, having all the data won’t do a lot of good.

3 Show the return on investment. Use the data to clearly show time savings, cost savings and any other benefits.

3 REASONS PRIVACY SHOULD BE TOP OF MIND

In May 2014, President Obama charged the Counselor to the President to lead a [90-day review](#) of data and privacy. The president wanted to know what new technologies were defining the data landscape and how big data and data analytics impact public policy and laws that govern privacy, particularly for citizens.

The findings from that review use the phrases “big data tools” and “data analytics,” but don’t get too caught up in the terminology. Yes, it’s important. But as we noted earlier, the terminology changes from person to person. What’s most important to understand is that data privacy doesn’t start or stop at a single point in time. Privacy has to be top of mind before the data is even created, and it must be maintained when the data is collected, stored, shared and analyzed.

As the report notes, “The opportunities presented by big data are considerable, but big data raises serious concerns about how we protect our privacy and other values.”

It’s true that “big data drives big benefits, from innovative businesses to new ways to treat diseases,” according to the report. But “the challenges to privacy arise because technologies collect so much data (e.g., from sensors in everything from phones to parking lots) and analyze them so efficiently (e.g., through data mining and other kinds of analytics) that it is possible to learn far more than most people had anticipated or can anticipate given continuing progress.”

Below are some critical privacy considerations from the big data report that you should heed on your big data and analytics journey:

1 **Big data and analytics tools can alter the balance of power between government and citizen.**

Government agencies can reap enormous benefits from using big data to improve service delivery or detect payment fraud. But government uses of big data also have the potential to chill the exercise of free speech or free association. As more data is collected, analyzed, and stored on both public and private systems, we must be vigilant in ensuring that balance is maintained between government and citizens, and revise our laws accordingly.

2 **Big data and analytics tools can reveal intimate personal details.**

One powerful big data technique involves merging multiple data sets, drawn from disparate sources, to reveal complex patterns. But this practice, sometimes known as “data fusion,” can also lead to the so-called mosaic effect, whereby personally identifiable information can be discerned even from ostensibly anonymized data. As big data becomes more widely used in the private sector to bring a wellspring of innovations and productivity, we must ensure that effective consumer privacy protections are in place to protect individuals.

3 **Big data and analytics tools could lead to discriminatory outcomes.**

As more decisions about our commercial and personal lives are determined by algorithms and automated processes, we must pay careful attention that big data does not systematically disadvantage certain groups, whether inadvertently or intentionally. We must prevent new modes of discrimination that some uses of big data may enable, particularly with regard to longstanding civil rights protections in housing, employment, and credit.

WHAT YOU CAN DO NOW: HOW TO CREATE A CULTURE FOR DATA ANALYTICS

Aside from budgets, cultural barriers pose one of the greatest challenges to introducing new technologies and new ways of thinking into any agency. There's a certain way that people have always done things, and sometimes learning a new way can cause angst or pushback.

So how do you foster an environment where people are open and excited about testing the waters and putting data analytics to use? Maybe you're well into your analytics journey, but you've hit a few snags and need some advice from your peers — or maybe you're just beginning that journey.

Wherever you are on that path, you can benefit from the lessons learned by others. We spoke with data experts at various government agencies to glean practical tips for creating a culture for data analytics.



Jeff Chen

Chief Data Scientist,
Department of Commerce

Learn the lingo.

When Chen was Director of Analytics for the New York City Fire Department, he rode along with firefighters to understand their daily experiences and struggles.

Tell the story about your data.

There are a lot of scientists at Commerce, and they have a lot of data. To help create interest in this data, the agency built tutorials that feature the scientists telling their data stories. "It's always a dialogue," Chen said.

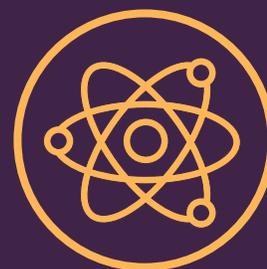


Jon Minkoff

Chief Data Officer,
Enforcement Bureau, Federal
Communications Commission

Brand your systems.

At the Enforcement Bureau, employees know their reports will be accepted only if they come from the bureau's audit tracking system. That means reports based on data from someone's Excel spreadsheet won't cut it. "It may be accurate, but it's not the root," Minkoff said. By demanding that all reports come from the bureau's system, it forces people to understand the system better, and it helps the bureau identify and repair gaps.



Kirk Borne

Data Scientist and former Professor
of Astrophysics and Computational
Science at George Mason University

Understand data as a literacy.

While technical skills are important, they are only truly valuable when combined with an understanding of data and problem-solving. "I always tell my students to understand data as a literacy," Borne said at a March 2015 GovLoop event. "Literacy is knowing how to use data, when to use it, what types there are, and what types of questions it can answer."

Recognize that data is an asset.

Doing so gives data a prominent role. "Data is a binary representation of real information, and real information is what we are trying to extract," Borne said. You can use data to create better insights, decisions, services to citizens or other agencies.



Caryl Brzymialkiewicz

Assistant Inspector General and Chief Data Officer at the Department of Health and Human Services Office of the Inspector General

Go on a listening tour.

What better way to find out what people are thinking than actually talking to them? They may not be down the hall, but follow Brzymialkiewicz's lead: She scheduled coffee meetings and lunches with people onsite and held Skype calls with people in the field. Ask them about their tension points, what they can't access but want to access, and what would help them do their jobs better.

Consider the talent you have on board.

Think about how to blend federal hiring with quick, commercial off-the-shelf software. Think about leveraging Presidential Management Fellows or a college intern to help you build your data analytics team. "Having one student on board can bring in new talent and thoughts," Brzymialkiewicz said.

Consider rotation programs.

Is there an entrepreneur in residence or an executive that you can bring on board to support your efforts?



Joshua Martin

Chief of Staff,
the Indiana Management and
Performance Hub

Make room at the table for a variety of participants.

The team at MPH is working with state agencies to explore what questions can be answered using their data and the state's business intelligence tool. State agencies are a part of the planning process and will help shape how the tool is used. "We really opened this up to the opportunity to continue to build on it with a number of different stakeholders and to bring everyone to the table," Martin said.

Pull from your past experiences.

Before starting his career as a state government employee, Martin spent six years in fire and EMS. His public safety background allows him to better communicate and relate to agencies that have a role in combating Indiana's opioid epidemic.

Keep everyone informed.

Martin and his colleagues have either met with or are scheduled to meet with every agency that is providing data to MPH. The goal is to show them what MPH is doing with their data and help them to better utilize their data to make decisions. "Once you can show your partners what you're doing and what value you can extract from their data, participation usually picks up fairly rapidly," Martin said.



Linda Powell

Chief Data Officer,
Consumer Financial Protection Bureau

Your data scientists and data management professionals have to understand the jargon of the business.

"If I walk into an economist's office and I don't understand him when he is talking about doing a linear regression, he is going to kick me out," Powell said. You have to understand the business and talk to them in their language, not in gigabytes and terabytes of data.

Understand what your agency customers or partners want to accomplish.

It's not as simple as walking into someone's office and asking them what they want. Sometimes they don't know, and it may be hard for them to articulate it. Here's where you need to sit down and have a two-way conversation. From that dialogue, you'll be surprised at how much you learn about their data analytics needs. (If you don't know what to ask, check out tips from Caryl Brzymialkiewicz. She provides some seed questions.)

CONCLUSION

Data is powerful. We can use it to shape policies; detect fraud, waste and abuse; and even save lives. But it takes more than data alone to drive better decision-making and ultimately better outcomes.

We also need the right tools to combine that data and search for patterns, anomalies and trends that otherwise would go undetected. Think back to the Michigan oncologist who intentionally misdiagnosed his patients: Imagine how many more lives could have been destroyed without the incriminating data to back up the whistleblower's claims.

We have to stop viewing data analytics as solely an IT project or something for people with computer science degrees. Some of the best data analytics teams are multidisciplinary. For example, the people on GSA's HR data analytics team come from diverse fields, with backgrounds in journalism, finance, and industrial and organizational psychology, to name a few.

As data analytics continues to permeate government, take the time to identify people who can add value to your analytics projects. Find the right people to train, and invest in them. You never know — your best data scientist may be a philosophy major. After all, they are good thinkers.

So what's next for data analytics in government?

Experts in the field expect to see greater use of data analytics to drive cybersecurity operations and predictions about future threats. When agencies are reactive — rather than proactive — the outcomes can be devastating. One recent example is the [massive data breach against Office of Personnel Management systems](#) that left millions of sensitive employee records exposed to hackers.

"Cyber is the No. 1 issue of our time," said Borne. "It's not just OPM, but so many instances where it would be better to predict something that doesn't happen than not to predict anything and deal with the fallout."

Some agencies are exploring next steps for using analytics to be proactive, rather than reactive, when it comes to security and carrying out their respective missions. Many of the agencies highlighted in our case studies are exploring those options, including the state of Indiana, GSA, the HHS Inspector General's Office and the Norcross Police Department.

Keep in mind that this resource guide is more than a compilation of government success stories. It's an invitation to reach out to the agencies you've read about, and a learning tool to enhance your data analytics journey.

ABOUT & ACKNOWLEDGMENTS

GovLoop's mission is to "connect government to improve government." We aim to inspire public-sector professionals by serving as the knowledge network for government. GovLoop connects more than 200,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C., with a team of dedicated professionals who share a commitment to connect and improve government.

For more information about this report, please reach out to info@govloop.com.

www.govloop.com | [@GovLoop](https://twitter.com/GovLoop)

Thank you to Information Builders, NetApp, Oracle, Pure Storage, Tegile, Teradata and ViON for their support of this valuable resource for public-sector professionals.

Author:

Nicole Blake Johnson, Technology Writer

Designers:

Jeff Ribeira, Creative Manager
Tommy Bowen, Graphic Designer
Kaitlyn Baker, Junior Designer
Martin Nera, Design Fellow



1152 15th St NW, Suite 800
Washington, DC 20005
Phone: (202) 407-7421 | Fax: (202) 407-7501
www.govloop.com
[@GovLoop](https://twitter.com/GovLoop)