

A *fresh* LOOK AT DATA

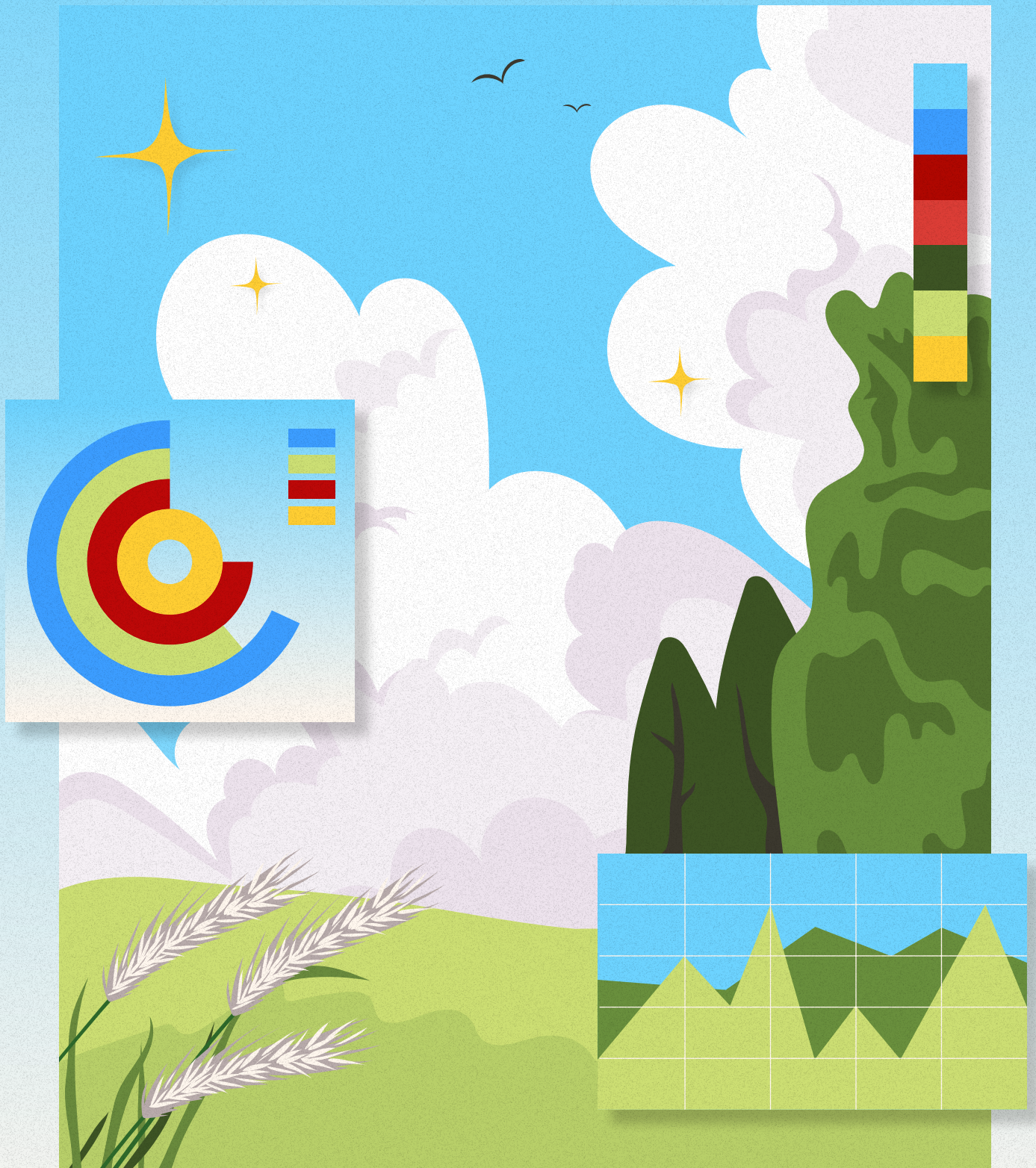


TABLE OF CONTENTS

✦ Introduction

PAGE 2

✦ Get Your Data House in Order

PAGE 3

✦ Where Data Makes a Difference

PAGE 6

✦ Using Data to Inform Policy

PAGE 10

✦ Conclusion

PAGE 13

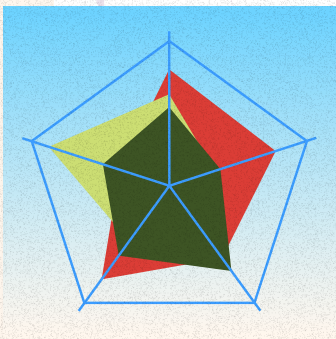


INTRODUCTION

Data has become an essential tool of government, driving how agencies serve customers, manage employees and track the programs' success. Increasingly, it serves as the basis for automating processes, using artificial intelligence (AI) and setting policy.

To take advantage of all that data can do, agencies must refresh their data strategies. They need to address governance in more systematic ways, investigate new use cases, and make the connection between policy and data analysis.

In this guide, we share some ways governments are updating their data approaches, and we hear from experts in the field. We hope you'll find incentives to revitalize data's role in your agency.



GET YOUR DATA HOUSE IN ORDER

Whether you're planning to use more AI or just want to improve analytics and tighten cybersecurity, good data management must be the foundation for your efforts. In 2024, agencies will need to get their data in shape to make the most of it.

The **Federal Data Strategy**, released in 2018, set high-level goals for using and handling data. Since then, agencies including the Department of Defense (DoD) have developed their own policies to implement those goals.

DoD based its 2020 **data strategy** on eight guiding principles that can serve as a useful model for other organizations in creating their own policies. Here are ways to carry out each principle.

Principle	Action Steps
Data is a strategic asset	- Leverage data for both immediate and lasting benefits. - Understand the value of your data.
Collective data stewardship	- Ensure accountability throughout the data life cycle. - Assign specific data management-related roles. - Set uniform security and privacy policies.
Data ethics	Consider potential individual/societal impacts in data collection, use and storage.
Design for compliance	Automate information management life cycle, data security and end-to-end records management.
Enterprise-wide data access and availability	- Promote interoperability and appropriate data access and sharing. - Adopt common data standards. - Make data understandable and put it in context.
Data fit for purpose	- Minimize sources of unintended bias. - Establish consistent practices for removing expired data. - Set quality standards and weed out low-quality data.
Data collection	- Collect data electronically when it's created. - Maintain an ongoing record of how and when the data is processed and revised.
Data for AI training	- Rethink data architecture to accommodate both structured and unstructured data. - Establish rules for tagging unstructured data to allow for efficient AI searches. - Label sensitive and private data to prevent inappropriate access.



According to [DoD](#), data should be:

Visible

Easy for users to locate.

Accessible

Simple to retrieve.

Understandable

Clear in terms of content, context and applicability.

Linked

Connected to related elements.

Trustworthy

Used with confidence.

Interoperable

Sharing a common representation.

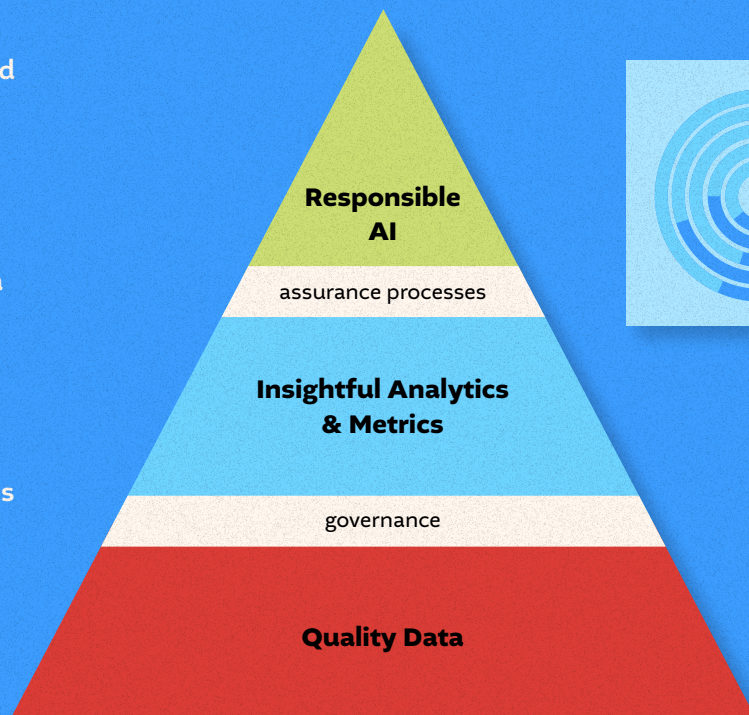
Secure

Protected from unauthorized use or manipulation.

The AI Hierarchy of Needs

Quality data lies at the foundation of thoughtful AI. To use AI responsibly, agencies should:

- ✦ Strengthen governance and remove policy barriers
- ✦ Deliver capabilities for enterprise business
- ✦ Improve foundational data management
- ✦ Invest in interoperable, federated infrastructure
- ✦ Advance the data, analytics and AI ecosystem
- ✦ Expand digital talent management



Source: [DoD](#)

Finding the Power of Data Through AI

An interview with Mark Hughes, Chief Strategy Officer; Stephen Moore, Chief Technology Officer; and Chris Steel, AI Practice Lead, AlphaSix

AI is changing the way government does business. Through automation, predictive analytics and advanced data analysis, AI is set to enhance operations and service delivery, said Chris Steel of AlphaSix, which provides data management platforms and tools for data analysis.

“Its transformative potential in government operations is vast, ranging from predictive analytics in defense and law enforcement to automating benefits processing, and everything in between,” he said.

And it’s time for agencies to get started, said Mark Hughes, also of AlphaSix.

“You don’t have to start large if people are interested in playing with the tools and seeing what they can do,” he said. “It easily scales as their needs scale.”

Getting Practical

Agencies are using AI for a range of applications, he said, including anticipating maintenance needs in public infrastructure, automating routine administrative tasks and using natural language processing for more responsive chatbots.

“Some of these innovations not only streamline processes,” Steel said, “but they make government services more accessible and responsive to citizen needs.”

AI also plays a significant role in cybersecurity. For instance, at a large federal agency, AlphaSix used machine learning (ML) and AI to create a complex anomaly-detection tool to protect sensitive customer data. “This not only secures critical data, but streamlines operations,” Steel explained.

Overcoming Hurdles

Preparing to adopt AI takes work, AlphaSix’s Stephen Moore said: “The first thing is just knowledge. You need [someone] who understands the technology behind the buzzwords.” Because there’s a shortage of skilled AI professionals, agencies may have to bring in vendors and consultants who can provide that expertise.

Another issue is having data that is AI-ready. “If your data’s spread all over the place in a bunch of stovepipe systems, it’s going to be hard to unlock the true value of looking at a higher level across datasets,” he explained.

Connecting all those silos unleashes vast amounts of data, often on systems that aren’t designed to handle those volumes, forcing analysts to summarize or use fewer datasets. “They’re missing the input that they would need to get the output, the analysis they’re looking for,” Moore said.

AlphaSix brings 13 years of experience in developing data tools and platforms to help agencies spend less time on data management and more on analytics, Hughes said.

Its solution is a data fabric, a system that starts with bringing in billions of records from diverse applications and making it available for analysis. The next step is to run queries and build the required analytics. “And the third phase is to add in AI and ML so you can start building models that can detect more in real time,” Moore said.

AI is ready for its role, Steel added. “It’s time to embrace AI as a transformative tool for government.”

WHERE DATA MAKES A DIFFERENCE

People often use vague language such as “vast troves” and “growing exponentially” to describe the amount of data at the government’s disposal, but those terms don’t say much. To get real understanding, consider some statistics:



By the end of 2025, the world will have generated, consumed, copied and stored **at least 180 billion zettabytes** of data. To put that in context, one zettabyte is the storage capacity of **250 billion DVDs**. It’s the same as 1 billion terabytes. And if you were to download all the data on the internet, it would take you **181 million years**.

The United States’ 5,381 data centers — more than 10 times the number in Germany, the country with the second most — handle much of that information. Every time we buy something online or post a YouTube video, we create more.

In the following pages, we look at innovative ways agencies use data to achieve their goals —specifically, to provide better services, develop effective workforces, measure government performance, and forecast and plan.

Data Storytelling

Facts are two-dimensional: They convey details but not causes or impacts. Yet if you put the data in context — if you tell its story — you make the details relevant to a wider audience. That leads to better decision-making.

After collecting and analyzing data, good data storytellers ask, “How can we maximize the value this data holds?” said **Megan Huss**, Director of the Process and Performance Improvement Program in the Office of Personnel Management’s **Center for Leadership Development**. What organizational or societal change can we shoot for?

Storytellers focus on the most critical insights, omit acronyms and jargon, tailor their stories to specific audiences, and guard against possible biases. They use colorful maps and other visuals and explain how the data affects people’s lives.

“One of the key reasons that [storytelling] is so effective is that we’re actually able to share not just what’s happening ... but we’re able to convey more about the why,” Huss said. “Until we ... make that linkage ... and encourage action from that data, the data really — in and of itself — would just have potential value.”

Customer Experience

To deliver consistently positive CX, agencies must identify pain points that stand in the way: the moments of confusion, delay, frustration or cyber risk that erode people's trust in government. Data helps agencies diagnose the root causes of those obstacles, whether, as the White House's **CX Executive Order** explains, they are statutory, regulatory, budgetary, technological or process-based.

Using Data

Agencies use a combination of **qualitative data** (what people think and feel) and **quantitative data** (what you can measure and count) to study CX. By analyzing post-transaction survey results, survey responses from employees and customer behavior data, such as average length of time to complete an online form, organizations can streamline service delivery and implement more **human-centered design**. Key elements of data-driven **CX** include:

- ★ Data sharing among agencies to promote consistency and save time.
- ★ Public-facing employees offering feedback on constituent experiences.
- ★ Accessibility redesigns based on empirical evidence and public input.
- ★ Officials using demographically representative data.
- ★ Agencies prioritizing reforms and celebrating small wins.
- ★ Understanding that technology is a vehicle to improve CX, not the goal itself.

Consider: Data “doesn’t give us answers, but it does give us breadcrumbs to follow, so we know where we might need to pay attention and in what way and what questions ... we need to start asking.”

— **Airis McCottry Gill**, former Executive Director for Employee Experience and Organizational Management, Veterans Experience Office, Department of Veterans Affairs

Workforce Management

Government struggles with recruitment and retention: Technical talent generally is hard to find, and industry employers lure qualified candidates with salaries and perks that agencies cannot afford. But data — collected regularly — helps agencies compete by enabling them to monitor trends and plan.

Using Data

Organizations are going beyond interviews and anecdotes, relying instead on quantitative data, such as that from anonymous monthly or quarterly surveys. The data feeds into an array of human resources (HR) functions, including hiring, onboarding, retention, career development, employee safety and time management. To foster a healthy and productive workforce, effective **data analysis**:

- ★ Integrates disparate HR systems and eliminates data silos.
- ★ Provides a holistic view of the agency workforce.
- ★ Continuously tracks employee engagement activity.
- ★ Identifies at-risk employees, skills gaps and training options.
- ★ Helps align staff and constituent demographics.
- ★ Helps managers track performance of hybrid/remote workers.

Consider: On average, it takes **more than 98 days** to fill a public-sector job, which is more than double the private-sector average.

— **Partnership for Public Service**

Performance Measurement

Good intentions can go awry and strategies get stale — so agencies must consistently measure, monitor and report on the progress of their initiatives and objectives. With concrete data, organizations can answer the question, “Are outcomes meeting expectations?” Then they can course-correct if needed.

Using Data

Agencies collect data at monthly or quarterly intervals, compiling several years’ worth of information (a long-term perspective is important), and establish benchmarks that offer context. Information is segregated into categories and subunits as appropriate (e.g., by field office) and presented in a **data dashboard** that tells a visual story. A **well-designed** performance measurement system, also called performance management:

- ✦ Draws a direct connection between collected data and the outcomes you want today.
- ✦ Breaks down silos and promotes internal collaboration.
- ✦ Is separate from the budget process, as much as possible.
- ✦ Values transparency and is thoughtful/strategic about public engagement.
- ✦ Distinguishes between key demographics and other such metrics.
- ✦ Focuses on outcomes, not outputs.

Consider: “I think about performance management being a tool to drive equity goals. It’s not about how many trees we planted or how many sidewalks we fixed — but where we’re making those investments.”

— **Rochelle Haynes**, Managing Director of What Works Cities

Planning and Prediction

Geospatial data, or geodata, is one of the government’s most exciting data opportunities. It helps agencies make evidence-based decisions — often using real-time information — regarding public safety, infrastructure, socioeconomics, disaster management, conservation, health care, national security and other concerns.

Using Data

There are **different types** of geodata, but the term generally refers to the location of specific things on the Earth’s surface, often derived from the **global positioning system** (GPS). Geodata feeds into **geographic information systems** (GIS) that analyze and display the data via online dashboards, for example. Agencies can track objects and events over time — such as the types and frequency of crime in specific neighborhoods — to identify trends. And the data can show live activity, such as vehicle movements. Agencies using geodata should keep the following in mind:

- ✦ Greater interagency collaboration reduces costly, duplicative data collection.
- ✦ Public/private partnerships have great data gathering and sharing potential.
- ✦ Cloud and AI technology offer powerful geodata storage and processing capabilities.
- ✦ Intellectual property agreements often restrict use of government geospatial data.
- ✦ Privacy, transparency, legal and security concerns exist.
- ✦ GIS skills are in demand and somewhat specialized.

Consider: Approximately 80% of all government data contains some reference to location.

— the International Research Centre’s **State of Open Data** project

How Search and AI Support Government Modernization

An interview with Dave Erickson, Public Sector Distinguished Architect, Elastic

Government wants to transition to cloud for its scalability and pay-as-you-go consumption model, among other benefits. But with some systems in the cloud and others on-premises, it gets hard to surface needed information.

As agencies modernize, they must be able to find, correlate and act on real-time data, regardless of location and format. In a hybrid cloud architecture, they must deliver search in a new way — one where “users don’t have to worry about data that’s distributed and disconnected sometimes,” said Dave Erickson, with Elastic. “We need to make it simple for users to ask the same question everywhere.”

Improved Data Availability

A partnership between Elastic and AWS is helping agencies to re-host and re-factor applications in the cloud and to re-architect their systems in support of improved information availability. “That partnership has been transformative,” Erickson said, effectively building “a big coalition of all the places where all the best data is, and making sure users can get to it.”

Elastic already empowers more than 100 federal civilian agencies in support of CISA’s Continuous Diagnostics and Mitigation (CDM) program. Cyberthreat hunters, for example, “can run federated queries across Amazon regions” when new threats emerge, Erickson said. With results returned in milliseconds, “people are transforming what was a week-long process before, into something that’s more at the speed of users’ attention spans.”

The ready availability of such capabilities in the AWS ecosystem helps government meet the challenge of doing more with less. “We want those in our GovCloud environments

to be able to buy from the marketplace, just like everybody else does,” Erickson said. As serverless technologies come increasingly to the fore, “we want to ensure that the cost curve for government stays predictable.”

The AI Factor

As agencies pursue modernization, they’re also wondering how artificial intelligence (AI) will factor in. Here, too, modernized search has a role to play.

Typically, generative AI-service providers want to feed AI models all necessary data, to ensure the most accurate outputs. But with privacy and other considerations, agencies want to control their data. “How do we give private data to AI, and stay in this realm of protecting privacy?” Erickson said.

The rise of the “vector database” can help resolve the issue. Such a database stores numerical representations of data, rather than the data itself, allowing AI models to securely access an agency’s proprietary data for context-rich, meaningful outputs.

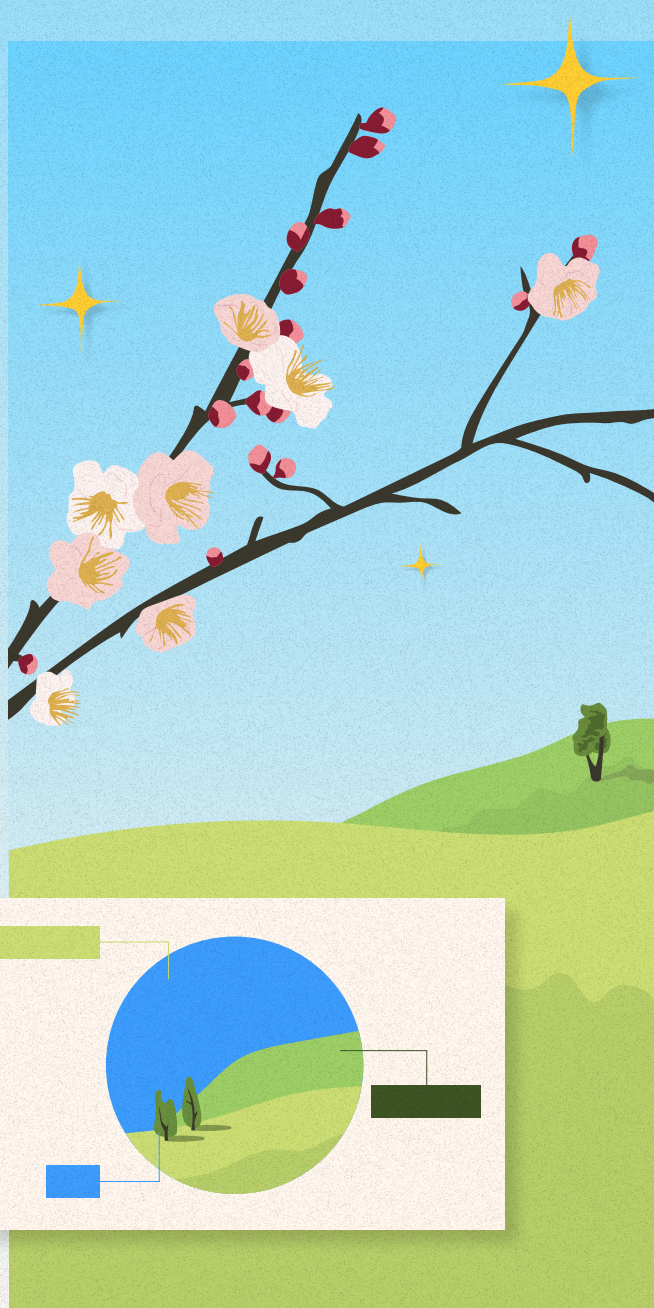
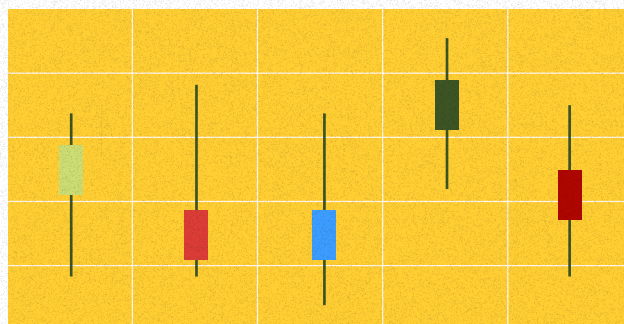
“A vector database [is] being used as the long-term memory for off-the-shelf AI models,” Erickson said. By keeping the vector in Elasticsearch, government’s private knowledge can be stored efficiently for rapid use.

That could help agencies use AI to support over-burdened workers. Today, “the person who is busiest is the person who’s getting questions all day: *Where do I find this? How do I do that?*” Erickson said. With AI, people will get sensible answers to plain-language questions. This promises “a world where every tool around me is really trying to help me,” he said.

USING DATA TO INFORM POLICY



An interview with Rachel Leventhal-Weiner, Director of Evaluation and Impact, Connecticut Office of Policy and Management



Governments collect and analyze a lot of data. But they don't always integrate it into the policymaking process. That can make programs less effective.

"Whoever makes decisions, having them see the value of using information is super important," said Rachel Leventhal-Weiner, Director of **Evaluation and Impact** at the Connecticut Office of Policy and Management. Collaboration between decision-makers and the people who work with data can be a challenge. Agencies need "to make that connection stronger [and] easier to manage, an easier bridge to cross," she said.

Everyone Is a Data Person

One way to facilitate that exchange is by shifting mindsets about who is a "data person." Often, people avoid using data because they think it's not relevant to their jobs. Leaders may be promoted because they are the best practitioners, rather than researchers, Leventhal-Weiner explained, so they may think data is someone else's department. It's not.

"When I teach people about how to use data, I start from the premise that everyone is a data person," she said. "We all use data in our personal lives. We use information all the time to make decisions about how to dress for the day [or] how to get from Point A to Point B, [but] there's some kind of disconnect between using information in our everyday lives and using information when it comes to our work."

Learning to Communicate

Another way to build collaboration is for data researchers and program staff to meet each other halfway.

“People who do data collection are not exactly sure what people want to know, so [they] just throw a lot of information out there and expect an audience to parse it for themselves,” Leventhal-Weiner said.

Meanwhile, “audiences typically do one of three things: They glom on to the first thing they see and [focus on] that one point. Or they read everything, but they feel confused. Or they see a lot of [data] and say, ‘I don’t want to deal with any of this!’ and make decisions without information,” she said.

Her advice for researchers is to think about questions the audience may have and present data that addresses them; for people on the program side, she recommends they formulate specific questions.

“Sometimes, the questions can almost seem too basic: How many people do we serve, or how are we doing this year compared to last year?” Leventhal-Weiner said. “Shrinking the universe to a question that people can answer, that’s really transformative.”

Using Data to Decide

A third approach to alleviating the data disconnect is to realize that information the agency already has can be applied to improving processes and rulemaking. “Sometimes things are right under your nose,” Leventhal-Weiner said.

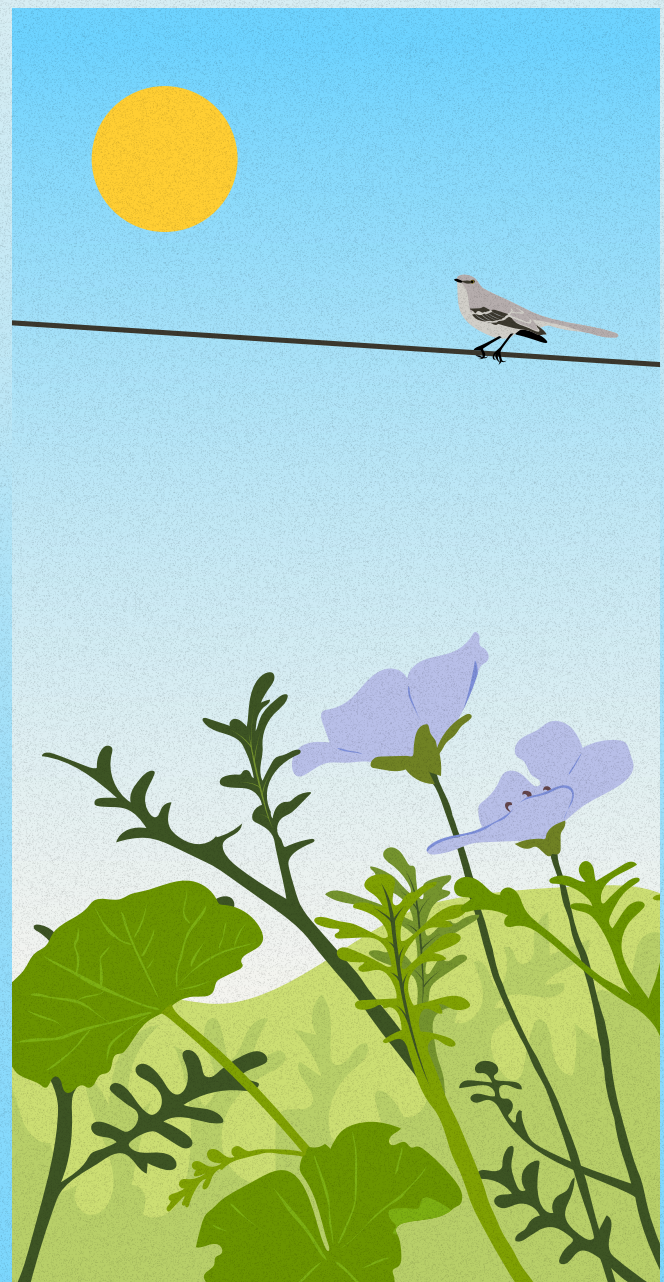
For example, if an agency has email addresses for all the people a program serves, it makes sense to send applications by email, rather than using a paper form. If most applications were returned within two weeks last year, you probably don’t need to leave the process open for 60 days this year.

In Connecticut, an Office of Early Childhood program provided in-home support visits for underserved families. The agency learned from a program evaluation that families enrolling before their children reached preschool age experienced better outcomes from the program. “So, we prioritized service for prenatal parents, rather than funding for everyone,” she said.

Using Data to Inspire

Leventhal-Weiner also recommended looking at your data through the lens of your mission. Develop measurements that can tell you how well you’re fulfilling it and what information could help meet it better.

“When you’re using your data routinely to look at what you’re doing – seeing if you’re meeting your mission – then you can start to dream about ‘what if?’,” she said. “Once you know the stats on your work, you can start to dream about what you want the stats to be.”



How to Extend Zero Trust to Data at the Edge

An interview with Red Hat's Michael Epley, Chief Architect and Security Strategist, and Kenny Peeples, Principal Cybersecurity Architect

With the rise of remote work and the explosion of the Internet of Things (IoT) generating large volumes of data in the field, agencies want to provide staff the ability to analyze and use that data in the field as well. This is the concept of edge computing. But there's a catch: They also need to secure that data.

"This is where zero-trust architecture (ZTA) comes into play," said Red Hat's Epley, "In this environment, it is critical to move from the castle-and-moat architecture of the past to address the needs of a hybrid cloud, and that includes the edge."

To extend ZTA to the edge environment, there are three key issues to consider.

Identity and Access Management

Identity and access management, while key to ZTA, present special challenges at the edge, according to Epley. An edge system typically is composed of many independent instances of systems, making it difficult to manage access privileges. Further, user credentials are at higher risk of misuse in the edge environment because there is less centralized control.

The solution? A distributed identity management solution. Such systems rely on multi-part approaches, such as blockchain (an immutable record of digital events) and Merkle trees, data structures that enhance the efficiency in blockchain. "There are a number of distributed identity systems being developed today, and they represent one approach to managing this problem," according to Epley.

IoT Fleet Management

Another challenge is the high volume and variety of IoT devices, which makes it difficult

to implement zero trust effectively, particularly as these devices operate according to automated routines.

Fleet-wide automation of a distributed identity and access management solution would help define allowable interactions and reduce risk across the entire system, according to Red Hat's Peeples. "With automation, Red Hat can implement the configuration-as-code across all those diverse devices, and be able to track the configuration drift."

Integration With Cloud Infrastructure

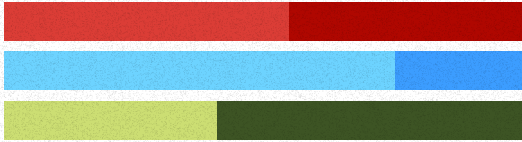
In extending ZTA to the edge, planners need to consider the role of cloud in supporting those widely deployed endpoints. That requires a different mindset.

"We have to stop thinking of edge and cloud and IoT as separate," Peeples said. "The practical reality is that these systems are going to be interoperating with each other. That means we will be making the zero-trust access-control decisions across our cloud infrastructure, as well as our IoT and edge systems."

And with edge data increasingly tied to the cloud, security efforts also will help ensure the validity of that vital information.

There's an urgent need to bring ZTA to the edge, where systems are subject to physical and network compromise to a much higher degree than non-edge-based systems. Applied at the edge, ZTA offers "a holistic security framework that addresses many pillars and capabilities of security" Peeples said.

WHAT'S AHEAD?



Where it once was seen as dry, difficult and the domain of specialists, data is now vital to how government serves the public and fulfills its mission. This guide is the first of three in 2024 that will explore ways agencies can use data now and in the future.

Thank you

Thank you to Alphasix, Elastic and Red Hat for their support of this valuable resource for public-sector professionals.

Content Team

Lauren Walker, Senior Staff Writer
Candace Thorson, Managing Editor
John Monroe, Director of Content
Susan Kirby-Smith, Senior Staff Writer
Adam Stone, Contributing Writer

Designers

Calista Lam, Graphic Designer
Andrew Blake, Motion Graphic Designer

About GovLoop

GovLoop's mission is to inspire public-sector professionals by serving as the knowledge network for government. Govloop connects more than 300,000 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 the public sector.

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

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



Click [here](#) to pre-register for the next two installments in the data guide series.