

STRENGTH IN NUMBERS: A BLUEPRINT FOR DATA ANALYTICS

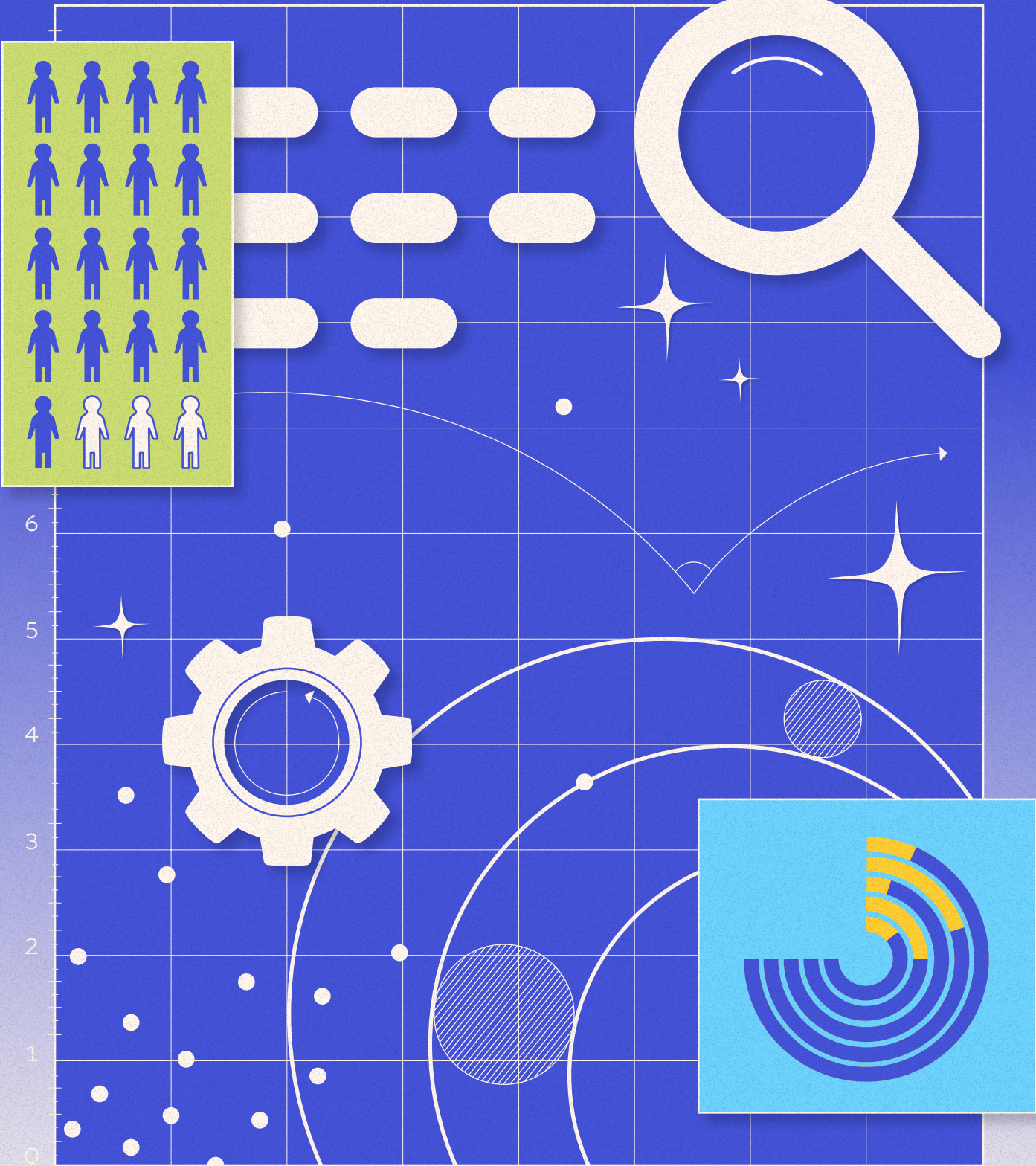


TABLE OF CONTENTS

✦ Introduction

PAGE 2

✦ Talk Data to Me

PAGE 3

✦ Where Data Takes You

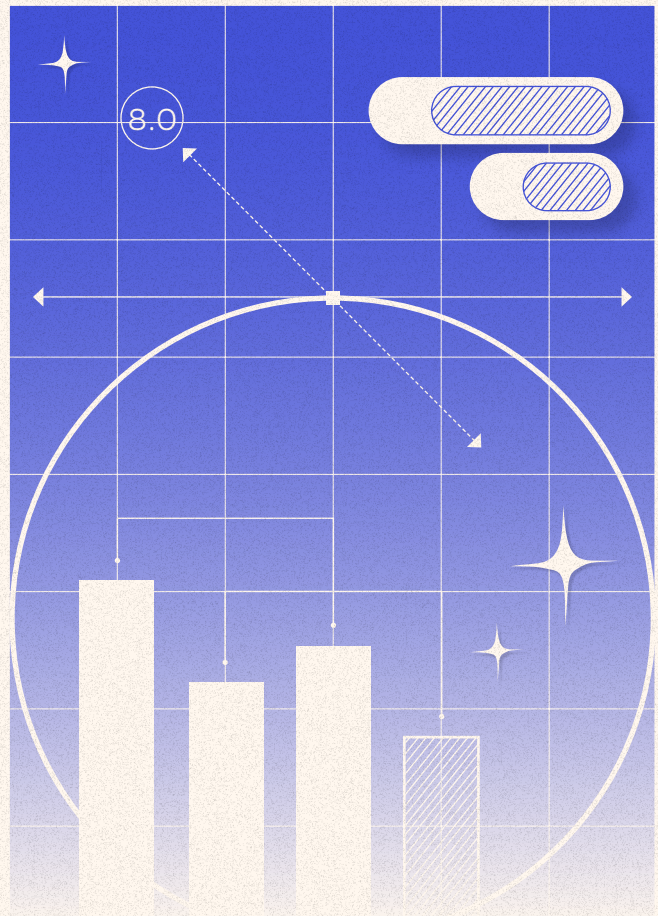
PAGE 6

✦ How to Use Data for Public Good

PAGE 10

✦ Conclusion

PAGE 13



INTRODUCTION

The first two installments of GovLoop’s 2024 data guide series explored the principles of effective data use and how to prepare your data for AI and other technologies. We looked at how to make your agency data-mature, why that’s important and what other organizations have achieved.

Now, in our final installment, we examine the motivations that drive agencies to collect, analyze and implement data initiatives. Facts and figures can improve service delivery, for instance, but sometimes that’s a side benefit. Sometimes a project’s original goal lies elsewhere.

Case studies in this guide highlight achievements related to each objective — e.g., policymaking, resource allocation, public service delivery, transparency and accountability, business development, and crisis management — and we dug deeply to find uniquely relevant examples for you.

The director of Georgetown University’s The Data Labs, which directly partners with government entities, explains how to use data to advance the public good. And industry experts discuss how technology can enhance government data initiatives, whatever their goals.

2024 was a busy time in the data world. The following pages will help ready you for 2025.

TALK DATA TO ME

Brush up on your data terminology.

Data Science vs. Data Analytics

According to the Harvard Business School, data science is “the process of building, cleaning, and structuring datasets to analyze and extract meaning.” Data analytics, on the other hand, refers to the process and practice of analyzing data to answer questions, extract insights, and identify trends.

Data Driven vs. Data Centric

A **data-driven organization** is not the same as a data-centric one. Agencies that are data-driven base their internal and external decisions on data, but the data still may be linked to individual applications or goals. The data is structured around the organization.

Data-centric organizations are structured around their data. Although this doesn't require a single database, it does require integrating data into all departments' operations, reducing silos and adopting a data architecture that's independent of specific uses, so it's available to all.

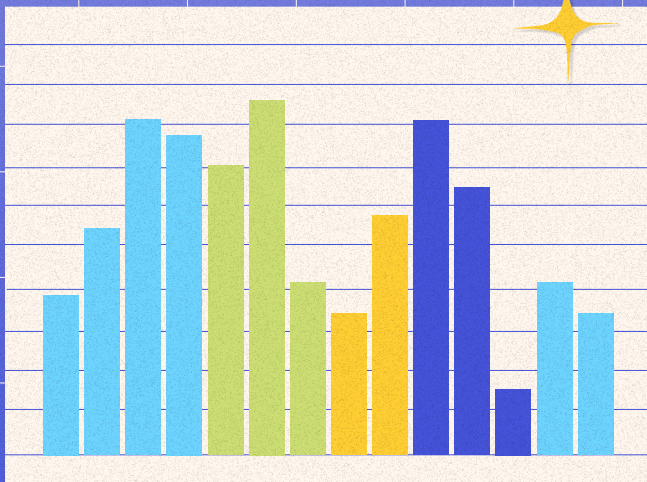
The 4 Types of Data Analytics

Descriptive analytics organizes raw data to show what's happened — for example, your monthly electricity bill — and indicate trends, such as whether you used more or less electricity in the past year.

Diagnostic analytics looks at possible reasons for those results. For example, it can look at corresponding trends, such as the average temperature or hours of darkness for the months in which your electricity use was highest, and may help you draw conclusions about power consumption.

Predictive analytics uses data to project what might happen in the future. For instance, if you use the most electricity when the days are shortest, you may be able to anticipate that you'll also use more power next winter.

Prescriptive analytics analyzes data to help determine next steps by pulling together known factors. In the electricity bill example, prescriptive analytics could use your electric use, work hours and the hours of daylight to suggest lowering your bill by adjusting the time you start work to take advantage of morning daylight.



Data Storage

Data warehouses: This is the most common type of data repository. It houses a tremendous amount of structured data, already cleaned and organized, that the organization uses for pre-determined purposes. This type of storage is often built on relational databases.

Data lakes: Unlike data warehouses, data lakes include raw data, both structured and unstructured – e.g., text, images, audio and video – often from a wide variety of sources. Because the data is unprocessed, it can be more difficult to work with than data in a data warehouse.

Data lakehouses: These platforms combine some features of data lakes – notably their ability to hold large amounts of data in different forms – with some of the organizing and analytic ease of data warehouses.

Data marts: You can think of these as smaller data warehouses for specific departments or functions.

Data fabric and data mesh: These are different ways of connecting the data held in various silos without having to move it into a centralized data warehouse or data lake.

Data Warehouse



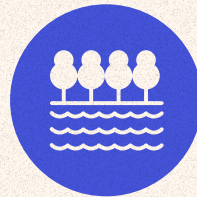
Reason for storing data is pre-defined

Data is processed and ready to be queried

Used by business professionals

Strong maturity model

Data Lake



Reason for storing data is undefined

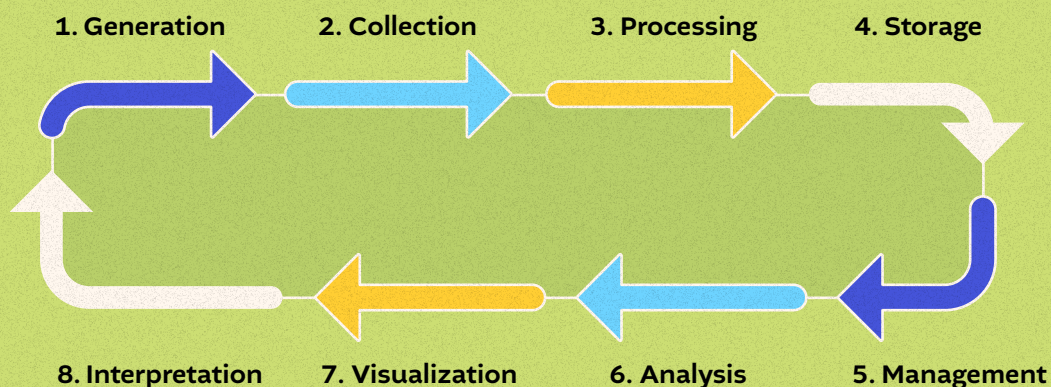
Data is left raw until it is needed

Used by data scientists

Emerging technology

Source: [What is a data lake?, Tech Target](#)

The Data Life Cycle



Source: [Harvard Business School Online, A Beginner's Guide to Data & Analytics](#)

HOW TO ENSURE YOUR DATA IS AI-READY



"Getting that data foundation in order often begins with understanding what data we have and what's contained in those datasets."

— Andrew Churchill, Qlik

"In the case of AI, unclean data can create a lot of problems — it can actually hinder the models."

— Piyush Bothra, AWS

[Watch the Video](#)

As agencies step up their efforts to leverage artificial intelligence (AI) and AI-based tools, they are quickly realizing the need to focus on data that will drive those solutions. Their challenge lies not in gathering enough data, but in ensuring their data consistently adheres to key governance practices, including data management, security and interoperability.

One area of particular concern is data hygiene. If an AI solution is based on a dataset that is outdated, incomplete or rife with errors, the outcomes or insights will be unreliable, said Piyush Bothra, Area Principal for Independent Software Vendors at Amazon Web Services.

Often, agencies need to begin by determining what data they have and what controls are required, said Andrew Churchill, Vice President for Public Sector at Qlik, because "you can't measure what you can't see."

In this video interview, Bothra and Churchill discuss steps that agencies can take to make their data AI-ready. Topics addressed include:

- The biggest barriers to getting data AI-ready
- Best practices in data hygiene to support AI initiatives
- Key criteria for assessing the readiness of your data

About AWS

Since 2006, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud. AWS has been continually expanding its services to support virtually any workload, and it now has more than 240 fully featured services for compute, storage, databases, networking, analytics, machine learning and artificial intelligence (AI), Internet of Things (IoT), mobile, security, hybrid, media, and application development, deployment, and management.

[Learn More](#)

About Qlik

Qlik converts complex data landscapes into actionable insights, driving strategic business outcomes. Our AI/ML tools, both practical and scalable, lead to better decisions, faster. We excel in data integration and governance, offering comprehensive solutions that work with diverse data sources. Intuitive analytics from Qlik uncover hidden patterns, empowering teams to address complex challenges and seize new opportunities.

[Learn More](#)

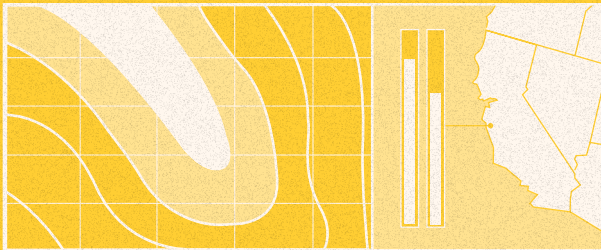


WHERE DATA TAKES YOU

Government uses data for several, often overlapping, purposes. Here's an overview of what those objectives entail, with links to further details, and case studies that show what's possible.

Policymaking

Agencies might identify a problem, but without data, they lack a clear, unbiased way to fix it. AI, machine learning and other analytics tools allow organizations to collect and swiftly analyze enormous datasets, which helps officials develop informed, equitable public policy, and predictive modeling helps organizations forecast outcomes. For example, sensors on streetlights can help officials understand traffic patterns and craft urban planning policies. Agencies need strong data governance, of course, to ensure that the data is trustworthy, valuable and ethically used.

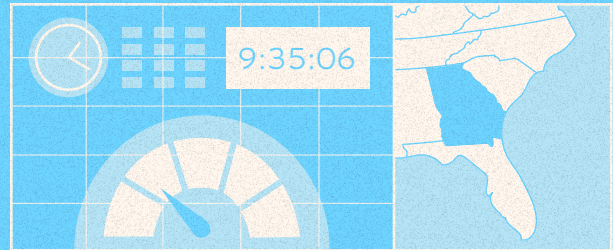


Case Study:

Salinas, California, had something in common with many cities larger than it: gang violence. Rather than tackle the problem alone, officials created a public/private Community Alliance for Safety and Peace (CASP). The group fast-tracks targeted services to young people by actively monitoring 24 major data points related to youth/gang violence prevention, intervention, enforcement and reentry, including police heat maps that show violent assault locations. The award-winning CASP also relies on a street outreach team that builds relationships with gang-affiliated youth. The heat maps have cooled down recently, and that reflects “thousands of youth not victimized by violence and thousands of families not living in fear,” said Jose Arreola, CASP Director and Salinas Community Safety Administrator.

Resource Allocation

In the public sector, funding and other resources are cherished and often uncertain commodities. That makes it all the more important to know where agency resources are going and how they ensure cost-effective use. Analytics tools help an organization identify budget-related trends and understand and resolve workforce challenges. With updated demographic data, government entities can decide how and where to dedicate their finite resources by asking questions such as: What populations are most at risk? What social services have the greatest reach? What programs are failing to meet expectations?

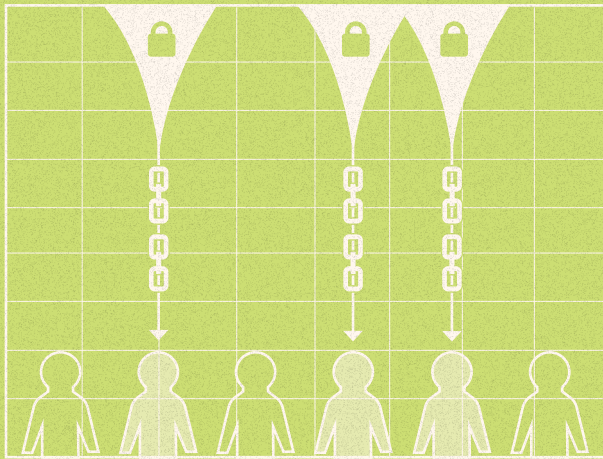


Case Study:

Like many public-sector entities, the Georgia Department of Corrections was concerned about staff burnout. So in 2022, in conjunction with its human resources technology vendor, the agency launched a pilot program that allowed prison and detention-facility managers to track their employees' working hours using data dashboards. Managers could see in real time how many hours individuals were working and make proactive scheduling changes before fatigue became a problem. “You can tell [supervisors] over and over again that they're working people too much, but until you show them in a way that jumps off of the page, it doesn't become clear,” said Cliff Hogan, the department's Director of Data Management and Research. “That's when people really start to say, ‘OK, I can see it now.’”

Public Service Delivery

In recent years, as virtual work became more common, government embraced data-centric technologies, such as cloud offerings and AI, and chief data officer roles. Those efforts help agencies dissolve data silos, collaborate with stakeholders, and improve service delivery and outcomes. By analyzing demographics and other trends, agencies can predict future obligations. In addition, digital technology now lets agencies personalize their services to individual constituents across various communication channels based on people's evolving needs and preferences.

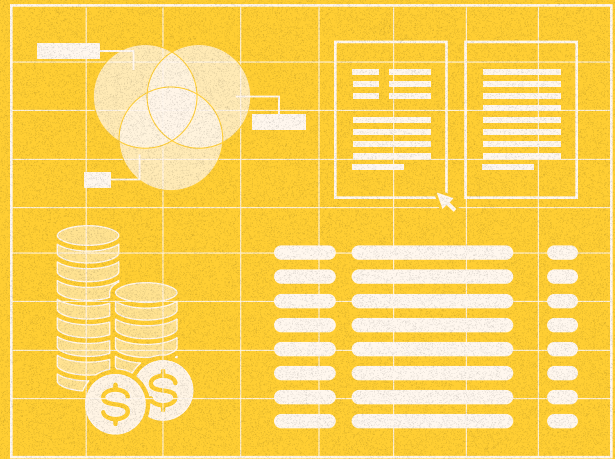


Case Study:

Many low-income families who are exempt from paying taxes can still benefit from tax credits designed to lift people out of poverty, but government often can't identify these families because they don't appear on tax rolls. To ensure that eligible residents receive funds they're entitled to, the California Policy Lab acts as a third-party facilitator in the state, linking two previously isolated state-agency datasets — regarding safety-net enrollment and state tax filings — through what's called "hashed linkage." Each agency encrypts its data before sending it to the lab. Families have received millions of tax-credit dollars as a result, and the linkage approach has potential for other programmatic challenges. "The data elements might be siloed, but the individuals behind the data are not," said Alissa Weiss, former Director of Government Innovation at Benefits Data Trust.

Transparency and Accountability

There's a lot of mis- and disinformation in the world today, and clear, publicly accessible data can counter it. Free, available data lets taxpayers track how their money is spent, which holds agencies accountable and clarifies often obscure government processes. Research, advocacy and policymaking all benefit from open data. Opportunities to innovate become clear. And online data portals, such as Data.gov and the Open Connecticut website, also reduce the number of public records requests.



Case Study:

Ten years ago, Congress passed the Digital Access and Transparency Act (DATA), which required, for the first time, the federal government to display all federal spending in one online location. Now, people can visit USAspending.gov to, for instance, find current data on federal contracts, grants and loans; explore fiscal details for individual agencies; and break down specific budget functions (Medicare, national defense, etc.). "Taxpayers had no reliable or consistent [way] to track how the monies were being spent," said Sen. Mark Warner (D-Va.), the bill's sponsor. "Passage of the DATA Act pushed us in a real way toward a more open and data-driven government."

Business Development

There's a direct connection between government data and private-sector job growth. For example, [McKinsey and Co. estimates](#) that open health data, by itself, adds more than \$300 billion to the economy each year. And some of the most well-known U.S. businesses, including [The Weather Channel](#) and [Zillow](#), were built using raw government data. They're examples of the private sector using public-sector facts and figures in ways that government either hadn't envisioned or couldn't implement alone. Sometimes, open data lures new industry to a community.

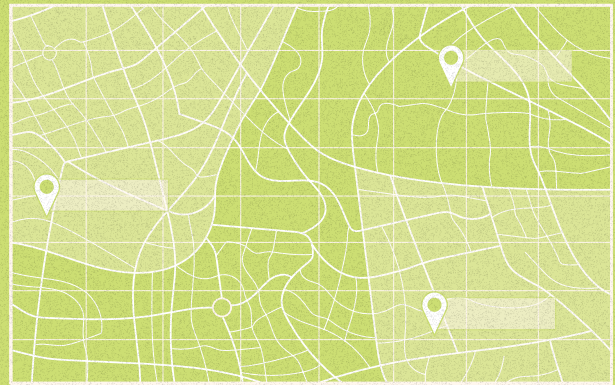


Case Study:

The East Tennessee Economic Development Agency (ETEDA) [website](#) tells the story of a region rich in business opportunity. In addition to colorful tourism photos, the site offers data regarding competitive business costs, technology access, workforce availability, commuting patterns, demographics, local services and more. It also includes statistics on tax rates and [relocation and expansion costs](#), detailed [economic overviews](#) for each of the region's 15 counties, and listings of available commercial properties. The approach is working, judging by all the [announcements](#) of new openings and expansions. [According to ETEDA](#), the goal is to promote a region that's "supported by a pro-business environment and talented workforce, driven by a diverse economy, and blessed with mild seasons and natural beauty."

Crisis Management

There are many examples of government using data, especially real-time statistics, to anticipate and deliver emergency services. Certainly, the [pandemic clarified](#) the need for local, state and federal data sharing; collaboration with private-sector stakeholders; and flexible IT that can consume, analyze and integrate masses of life-altering facts and figures. [AI and other automation](#), cloud technology, and similar data-focused tools are essential, as are data governance, security, and ethical standards.



Case Study:

The U.S. Census Bureau's online [Emergency Management Hub](#) offers local data for nearly every type of natural disaster in the country and for man-made events such as Baltimore's Francis Scott Key Bridge collapse. The hub's [My Community Explorer](#) tool, for example, is a flexibly designed, map-style dashboard that shows relationships between underserved areas and natural disasters. Demographic, economic and other data identify how many residents are disabled, without cars, or live alone; unemployment rates and average incomes; languages spoken at home; and broadband access, among other details. The website also offers [resiliency estimates](#) for individual communities. Taken together, the technology promotes understanding of "the demographic and economic characteristics of hard-hit areas to help ensure more targeted and equitable response and recovery efforts," [said](#) the Bureau's Deputy Director, Dr. Ron Jarmin.

THE EMERGENCE OF AN AI FACTORY MODEL



"If you're a data domain expert now, you don't necessarily need to be an [machine learning] expert to start building ML models. And that democratizes it, which means the people who know the mission and the data best can now be the ones initially coming in and saying, 'Hey, this would work.'"

— Winston Chang, Snowflake

Watch the Video

Across the federal government, agencies are exploring how they can use artificial intelligence (AI) to solve long-standing problems and discover innovative ways to manage operations and deliver services. By necessity, most agencies have focused on tactical objectives, limiting their risk while they learn about the technology and build out their use cases.

Their challenge now is to lay the foundation for more strategic initiatives by adopting a set of repeated practices that produce consistent results, making it easier to scale those operations. In this video interview, Winston Chang, Chief Technology Officer for Global Public Sector at Snowflake, describes what he calls the AI factory model.

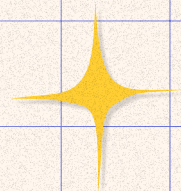
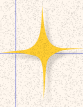
Topics addressed include:

- The three phases of an AI factory model
- How this approach makes it possible for non-experts to start building models
- The importance of incorporating interdisciplinary teams in the process

About Snowflake

Snowflake for Public Sector makes it easy to modernize and accelerate cloud migration, share data securely, and ensure data governance for improved resiliency and enhanced mission outcomes. Snowflake's AI Data Cloud enables organizations to enable a data-informed government. Leverage the Snowflake AI Data Cloud and platform to power IT modernization, improve citizen services, and drive innovation and efficiency. Snowflake delivers a FedRAMP-authorized, seamless experience across multiple public clouds.

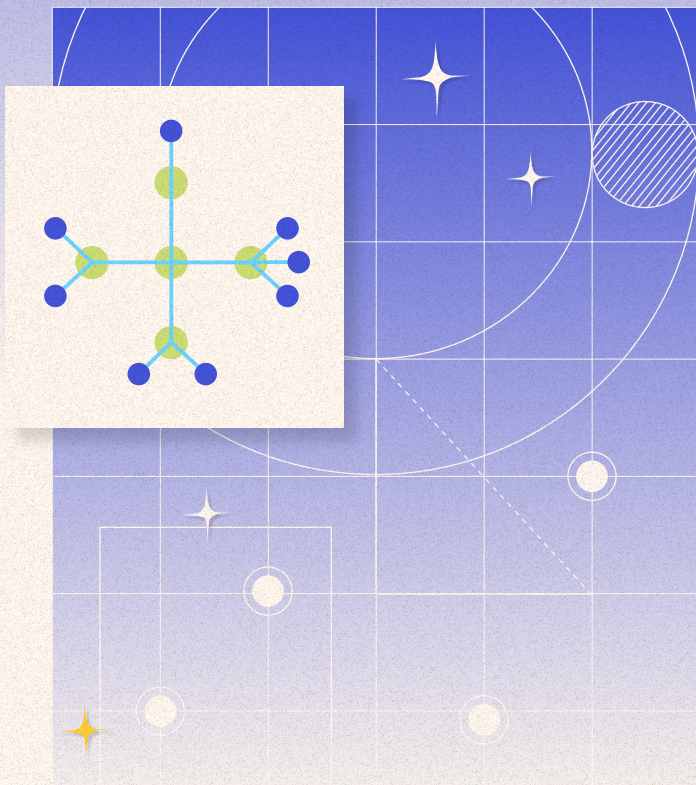
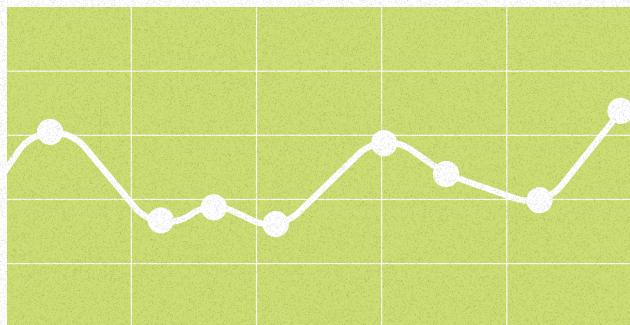
Learn More



HOW TO USE DATA FOR PUBLIC GOOD



An interview with Ali Benson,
Director of Data Labs, Beeck Center
for Social Impact and Innovation,
Georgetown University



Most of us are familiar with projects gone wrong, with indecision, delays and wasteful sidetracks when no one understood the project’s purpose. But up-front analysis can avoid those speedbumps, and it’s especially vital for data-driven projects, such as determining how an agency feeds residents experiencing food insecurity.

In other words, “you want to make sure that you’re solving the right problems, first and foremost,” said Ali Benson, Director of the Beeck Center for Social Impact and Innovation’s Data Labs. To do that, define the challenge and then “ask yourself, ‘OK, what data do we have, what data do we need and what are potential solutions for it to solve this very specific problem,’” Benson said.


“You may be trying to boil the ocean,” she added, “when really you can take a targeted population that may be disproportionately experiencing challenges and you can focus on that, achieve a win, demonstrate what’s possible and then scale from there.... Up-front work will help prevent you from getting further down the road and then realizing, ‘Oh shoot, this wasn’t feasible.’”

“Up-front work will help prevent you from getting further down the road and then realizing, ‘Oh shoot, this wasn’t feasible.’”

— Ali Benson

Play-by-Play

There’s guidance to help you make thoughtful decisions early on. The award-winning [Data Labs program](#) uses human-centered design to help states launch data-sharing projects that improve public services and people’s lives — though its tactics apply to all levels of government, whether an agency is expanding an existing project or starting a new one. The [Data Labs Playbook](#) offers detailed best practices and dozens of worksheets, checklists and other supplementary resources to guide you.



For instance, the playbook explores the importance of data governance and gives advice on remediating potential bias. “If you’re [examining] the work to identify where that bias exists and documenting it and looking for ways to mitigate it,” Benson said, “that ultimately will help ensure that your work is as equitable as possible, and as accurate.”

Data initiatives should be team efforts, she added. “You should be convening a cross-functional team of doers and implementers, of people in the governor’s office, data experts, etc.,” said Benson. And leverage your CDO or other executive-level data official because they have an “enterprise lens” and connections to funding streams and key contacts.

Tell a Story

Even the best data is two-dimensional, so tell your project’s story to build support for it, she advised. “Take the time to figure out, who are my key stakeholders, what do they care about, and how can I draft a compelling narrative to help them understand why this work is important [and] how they can play a role in moving it forward.” Ask yourself, whose life changes if this project goes well?

In Colorado, the Department of Local Affairs’ (DOLA) Division of Housing knew who needed help: thousands of “chronically unhoused” Colorado residents — a demographic that swelled 266% between 2007 and 2021, more than any other state’s. Working with Data Labs, DOLA developed a proposed pilot program to increase access to affordable dwellings.

DOLA officials drafted an action plan for their work and pitched it to leaders, and within six months, the department launched the program, Benson said. The results were very positive. By 12 months, the department secured a historic level of funding — \$250 million — that included money for data initiatives and built on the pilot’s success. States rarely fund data projects, Benson noted.

Colorado eventually kicked off a second pilot program 18 months after the first one using lessons learned from the initial effort. The state is a perfect example, said Benson, of “launching a pilot that secures an easy win, demonstrates what’s possible and then lays the groundwork for ... efforts in the future.”

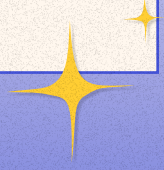
Working with data brings technical challenges, of course: bridging data silos, dealing with legacy IT, addressing skills gaps, securing long-term funding, and others. Certainly, “information exists in disparate places, and they don’t really talk to each other, yet policy areas are so complex and interwoven,” Benson said.

But effective data projects also require something less substantive, she believes: “This work moves at the speed of trust.”

Defining the Problem: Questions to Ask

- What is the current problem as you see it? What do you know or not know about it?
- What questions do you have about the problem, the issue area and your specific project? What do you need to answer these questions?
- Has any other government (local, state or federal) encountered a similar concern? What can you learn from what they tried?
- What legislative priorities should you be aware of? Who leads them?
- Who are the key stakeholders? Allies? Legal resources? “Make or break” stakeholders?
- Did stakeholder conversations identify common themes? Issues? Challenges? Obstacles?

Tip: Ask yourself, whose life changes if this project goes well?



DATA MANAGEMENT'S SPECIAL INGREDIENT: BACKUP AND RECOVERY



Sean Applegate

Chief Technology Officer, Swish

Watch the Video

Managing data is no small feat. An array of data formats — from video files to Microsoft Excel spreadsheets — make it difficult. Then add legacy technology, data and communication silos, and limited-purpose applications to the mix. There's ransomware, too, and the public sector is a prime target. So to use data securely, as a strategic asset, agencies need a simplified, cohesive approach to cyber resiliency.

For instance, an AI-driven backup and recovery system helps you protect, extract and contextualize data quickly from either yesterday or years back. And teamwork makes implementation more successful.

In this video interview, Marlin McFate, Cohesity's Field Chief Information Security Officer and Chief Technology Officer for Public Sector, and Sean Applegate, Chief Technology Officer with Swish Data, provide insights into how agencies can secure and maximize their data's potential.

Topics addressed include:

- Agencies' data management challenges and opportunities
- What a thoughtful backup and recovery system can achieve
- How data management relates to the overall modernization landscape

"Think about the things that you want to accomplish that data management or data modernization would require. And don't just think on the infrastructure side [or] on the AI side.... It really kind of spans the gamut."

— Marlin McFate, Cohesity

"Teamwork is the name of the game, and ... agencies that are able to do that easily, cost effectively, and then unlock value from it are going to be the standouts in the next 10 to 15 years."

— Sean Applegate, Swish Data

About Cohesity

Cohesity is a leader in AI-powered data security and management and empowers the federal government to use one platform with limitless scale to protect, secure, and manage data seamlessly — anywhere the mission requires.

Learn More

About Swish Data

Swish is a provider of technology solutions and engineering services to the U.S. federal government with a focus on high-quality outcomes for customers. Since 2006, Swish has delivered high-performance solutions and services to the federal government market ensuring that customer's digital service capabilities, performance, and security exceed expectations and requirements. Swish is a Service-Disabled, Veteran-Owned and HUBZone-certified Small Business.

Learn More

CONCLUSION

This guide is Part 3 of our 2024 data guide series. Here's a window into the first and second installments.

Part 1: A Fresh Look at Data

Get Your Data House in Order

Learn guiding principles and strategies for maximizing your data's potential.

Where Data Makes a Difference

Read specific tips on using data for customer experience (CX) improvements, workforce management, performance measurement, and planning and prediction.

Using Data to Inform Policy

Hear about the tactics that one state official adopts when using data to facilitate policymaking.

Part 2: Leveling Up With Data

A Grown-Up Look at Data

Explore what "data maturity" means and why agencies should achieve it.

Government Case Studies

Learn how organizations use data to improve health care delivery, make procurement easier and more effective, guide emergency response, update employee performance requirements, re-envision road maintenance, and address water shortages.

Data Governance: How to Use Data Effectively

Read about data governance best practices from an in-the-weeds government official.

Thank you

Thank you to Cohesity/Swish Data, Qlik/AWS and Snowflake for their support of this valuable resource for public-sector professionals.

Content Team

Candace Thorson, Managing Editor

John Monroe, Director of Content

Lauren Walker, Senior Staff 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)

