Decision Intelligence:

New Possibilities for Data-Based Decision-Making

A GOVLOOP GUIDE



Introduction

Agencies have spent the last several years establishing a foundation for data-based decision-making. They've developed data strategies and policies and built out agency infrastructure, among other endeavors.

The task now is to extend those capabilities and use data in new, more productive ways. This guide explores such options and makes a concerted effort to focus on data initiatives with tangible impacts. We look at homelessness, opioid abuse, climate change, and societal equity, to name just a few concerns, and provide concrete ways to affect change at your agency.

We also examine data privacy and ethics and offer options for cultivating a data-literate workforce. And since data and technology are closely linked, we hear from industry experts regarding the technology that facilitates informed decision-making.



Table of Contents

4 THE BASICS

5 CONSIDER THIS

6 DATA PRIVACY

8 DATA ETHICS

9 EMPLOYEE SKILLS

12 REAL-TIME DATA

16 OPEN DATA

20 DATA EQUITY

25 CITIZEN SCIENCE

30 ADDITIONAL RESOURCES

31 TAKEAWAYS

The Basics

Quantitative data is numbers-based. It's countable or measurable, and it tells us things such as how many, how much or how often.

Qualitative data is interpretations-based. It's descriptive, relates to language and helps us understand why, how and what happened to cause certain behaviors.

Which Types of Data Collection Are People Most Aware Of?

- Background Check
- Sexual Preference
- Government ID
- Tax Information
- Gender
- Audio/Voice Recording
- Internet Service Provider
- Address
- Device Address/IP
- Browser Activity
- Phone Number
- Payment Information
- Contacts
- Camera/Photo Access
- Location

Source: Security.org

Data Collection Methods

1. Forms & Questionnaires
2. Interview
3. Observation
4. Documents & Records
5. Focus Groups
6. Oral Histories
7. Combination Research
8. Online Tracking
9. Online Marketing Analytics
10. Social Media Monitoring

Source: Global Patron

The Benefits of Interoperable and Connected Government Data



Improve residents' experience of public services



Increase administrative efficiency



Enhance data protection and privacy



Deliver the value of open data



Enable data-driven policy making



Reduce fraud, waste, and abuse

Source: McKinsey & Co.

Data Literacy Skills



Thinking Critically about information yielded by data analysis.



Understanding data analytics tools, methods and when, where to use them.



Interpreting data visualizations, such as graphs and charts.



Recognizing when data is misrepresented or used misleadingly.



Knowing what data is appropriate to use for a particular purpose.



Communicating information about data to people lacking data literacy (aka data storytelling).

Source: TechTarget

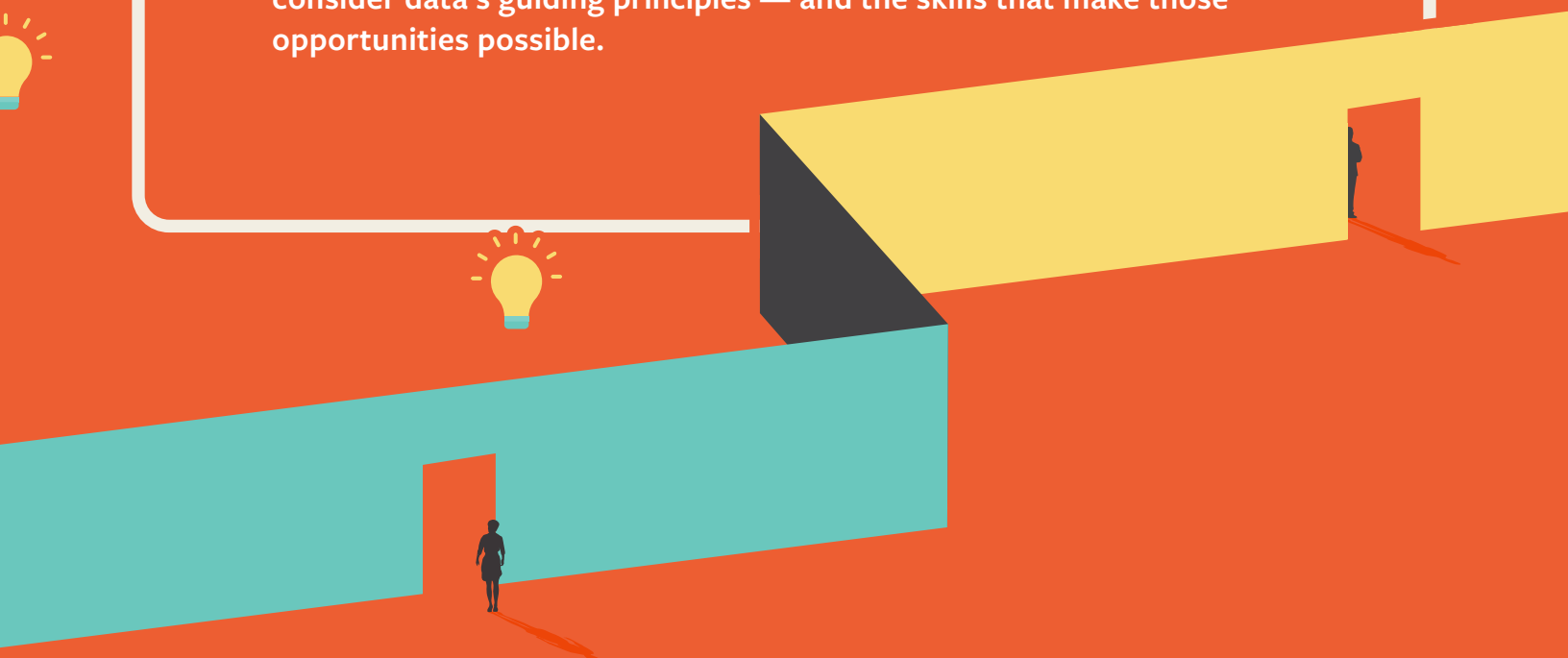
5

Consider This

Data has many sources, and much of it is personal: our Social Security numbers, financial information, contact details, and more. That data can say much about us, but where should government agencies draw the line? At what point do they have a moral imperative to stop collecting and using our personal information?

And beyond that, what do agency employees know about collecting, analyzing and caring for the data at their disposal? How can public-sector staff become data literate?

So, before we explore specific data opportunities, we must consider data's guiding principles — and the skills that make those opportunities possible.



Data Privacy

Government agencies have masses of personally identifiable information (PII), such as Social Security numbers, health records and financial details, at their disposal, and they need effective ways to gather, manage, store and secure it. But to what extent should organizations collect personal data, and for what purposes? And to whom does it belong?

Data privacy is about an individual's right to have some control over how their personal information is collected and used. As the world becomes more technologically advanced and government relies on more automated systems, our personal information can go well beyond its original source — and we might have little idea what happened to it.

For instance, many agencies are using automated decision-making (ADM) systems. They can be very helpful: Using artificial intelligence (AI), ADM systems assist or replace human efforts to make potentially life-altering decisions related to public benefits, policing, housing, health care, immigration, education, emergency response, child welfare and other concerns.

However, it's often unclear how much data an ADM system uses and what it ultimately does with the information. Does it eventually sell your personal information to a partner, parent company or subsidiary?

The federal government has some laws related to sector-specific data privacy. The U.S. **Privacy Act** establishes rights and restrictions regarding the data that government organizations hold. The **Health Insurance Portability and Accountability Act**, commonly known as HIPAA, protects health care and health insurance personal data. **The Family Educational Rights and Privacy Act** safeguards student records.

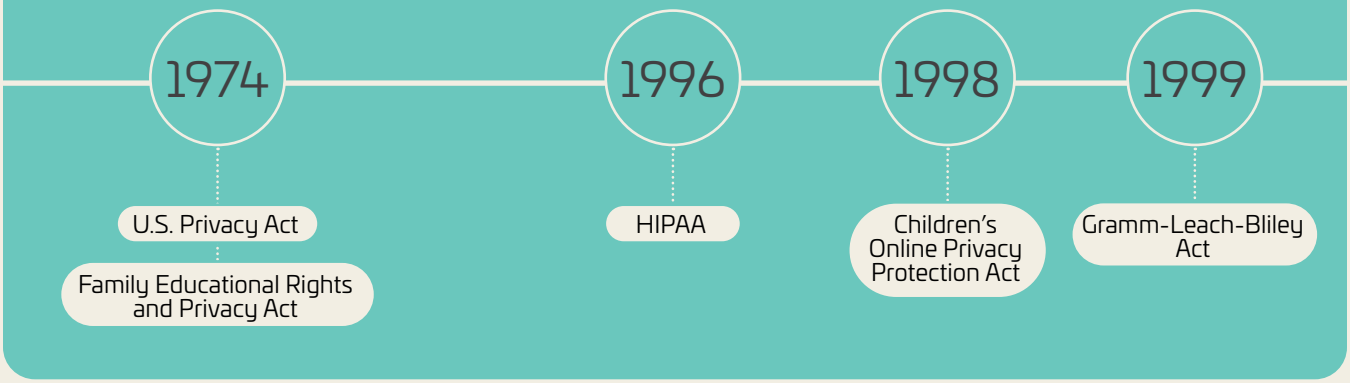
And the **Gramm-Leach-Bliley Act** protects personal, non-public financial information, while the **Children's Online Privacy Protection Act** regulates the collection of personal details for children younger than 13. Congress has considered, but not enacted, broader privacy legislation.

At the state level, California enacted the first state privacy law in 2018 and, via ballot initiative, established a more comprehensive version in 2022. Four other states — Colorado, Connecticut, Utah and Virginia — enacted statutes that take effect in 2023. By and large, all five laws differ from federal requirements in one important way: They establish a person's individual right to control their personal information — akin to the European Union's expansive General Data Protection Regulation — while federal laws simply take a "harms-prevention-based" approach that tries to prevent or mitigate potential problems.

The American populous has little faith in public and private organizations' ability to protect personal data, and people struggle to control their own information. According to a recent survey, 70% of U.S. respondents think it's become harder to control who accesses their personal, online information. And only 26% of respondents think that current data privacy laws are good enough.



Timeline of Key Federal Privacy Laws



What People Think About Data Privacy

A recent survey asked U.S. residents about their data privacy thoughts and concerns. Below are some of the findings.

Q: Do you agree or disagree with the following statements?



Source: Ipsos

What Does the U.S. Privacy Act Do?

The 1974 Privacy Act relates to the personal information government agencies hold. Among other provisions, it gives individuals the rights to:

- Request access to their personal data and request corrections, if necessary.
- Know how agencies use their personal data after collecting it.
- Access the data (on a person-by-person basis) based on the person's role in a company; in other words, someone in a non-governmental organization can view a constituent's personal data, but only the kind (e.g., financial) that specifically relates to the person's role at the company.

Data Ethics

By definition, data ethics refers to the norms of behavior that promote appropriate judgments and accountability when acquiring, managing or using data. There are three goals: protect civil liberties, minimize risks to individuals and society, and maximize the public good.

It is about more than complying with legal requirements; in fact, some of the most controversial realms, such as AI, are largely unregulated anyway.

But AI's expansive use has made data ethics increasingly important. Think about all the constituent data that AI-driven ADM systems use when they make decisions — on governments' behalf — regarding public benefits, emergency response, child welfare, and other concerns. Or about facial recognition systems that are less accurate with darker skin tones.

According to the Federal Data Strategy's Data Ethics Framework, "The ethical use of data is both an individual and organizational responsibility, involving many stakeholder groups that can be internal or external to an agency."

"Instead of looking at issues from a single perspective," the Framework cautions, "ethical decision making is best achieved by taking a holistic approach and widening the context to weigh the greater implications of data use."

4 Basic Principles of Data Ethics

OWNERSHIP



Individuals own their data or their information. They have the control and right to decide with whom they share it and at what levels of access.

TRANSPARENCY



Data should be acquired with consent from data subjects who should be explicitly informed where their data is being stored, for how long and the purpose of accessing it.

INTENTION



Data should only be used for its intended purpose and nothing more. Ensure it is being used or shared fairly and honestly with authorisation.

PRIVACY



Data should be secured safely to ensure there's no personal identifying info that runs the risk of identification of individuals or groups. Precautions are taken to prevent data leaks/breaches & dissemination.

Source: Medium

Significant AI Ethical Issues



Source: Deloitte

7 Data Ethics Tenets

1. Promote Transparency
2. Respect Privacy and Confidentiality
3. Act with Honesty, Integrity, and Humility
4. Hold Oneself and Others Accountable
5. Uphold Applicable Statutes, Regulations, Professional Practices, and Ethical Standards
6. Respect the Public, Individuals, and Communities
7. Stay Informed of Developments in the Fields of Data Management and Data Science

Source: Data Ethics Framework

Employee Skills

How Texas Is Cultivating a Data-Savvy Workforce



An interview with Neil Cooke, Texas' Chief Data Officer

Having worked in Texas state government for more than 15 years, Neil Cooke understands firsthand the difference that data can make.

Cooke began working for Texas in 2007, when he took a job at the Department of Agriculture, first as a systems analyst and later as a management analyst and project manager. During his seven years there, he and his team looked at how to use data to measure various programs' performance.

In the process, they began to apply data management principles, although that term was not yet in vogue. "We were asking questions around the completeness and accuracy of the data, and about getting it where we needed it and when we needed it," Cooke said.

The effort paid off. For one program, they increased service delivery by 20% — and without affecting the quality of service or the size of the staff.

He did similar work on a larger scale when he was Director of Performance Management at the Texas Health and Human Services Commission, helping to design an enterprisewide performance management system supporting hundreds of programs.

Now, Cooke, who was appointed Chief Data Officer (CDO) in December 2022 after two-plus years as a data officer, focuses on helping other state employees learn how to bring data to bear on their work. "The goal is to instill that data-focused philosophy throughout the organization, so that it's not limited to the data team or IT department," he said.

The Building Blocks of Data Literacy

In January 2022, Texas put together an interagency working group to develop an approach to improving the state workforce's data literacy. The learning path for data scientists is well understood, but what does it look like for people who want to be savvy consumers of data?

"A data-savvy workforce understands how to collect and analyze data but also understands its significance," Cooke said. "They use data to inform their decisions, develop strategic plans and measurably contribute to the success of the organization."

One key concern is data lineage: Where did the data come from? And how complete, accurate and timely is it? "Accurate and timely decisions depend on employees knowing how to read, understand, work with and communicate with data," Cooke said.

The working group identified six core competencies of data literacy:

- **Understand:** What is data and its impact?
- **Find/obtain:** Where can I find the data I need?
- **Read:** How do I interpret it and what should I know?
- **Manage:** What are the principles of data management?
- **Using:** How is data prepared and how do I analyze it?
- **Communicate:** How do I talk about data with others?

But before they could begin building a data literacy program, the group needed to establish a baseline. So, they asked its members, who represented more than 20 agencies, to assess their respective teams' data literacy levels.

The working group then conducted a gap analysis, looking at the current state of data literacy and where the group wanted to be. With that analysis in hand, it put together a plan to close that gap, Cooke said.

A Three-Pronged Approach

The first step was creation of the Texas Data Literacy Program. It was designed to give employees across state government and higher education the chance to improve their data competency through a series of courses that cover a range of data-related topics.

The program starts with 10 foundational courses applicable to everyone. They cover topics such as:

- Introduction to data literacy
- History of data management and the Office of the CDO
- Data governance and the data life cycle
- Data quality
- Data privacy and ethics
- Data security
- Data sharing
- Roles and responsibilities
- Storytelling with data
- Records management

Cooke said the state envisions a three-pronged approach to offering the training.

- **Central learning management system.** The courses are hosted on a learning management platform. Employees of a state agency or institution of higher education can request access through the Texas Data Literacy Program site.
- **YouTube.** Some courses also are available through the Texas Department of Information Resources' YouTube channel. "We believe it is important to reach a larger audience of data users who might not necessarily be interested in earning certificates of completion," Cooke said.
- **Individual agency programs.** In the future, they also might allow individual agencies to host courses on their own learning platforms, so they can administer the training to their staff in a way that works for them.

A Work in Progress

Several of the initial foundational courses already are available, with others in development. Once the foundational courses are complete, the program will produce courses that offer more advanced training on some of those topics, Cooke said.

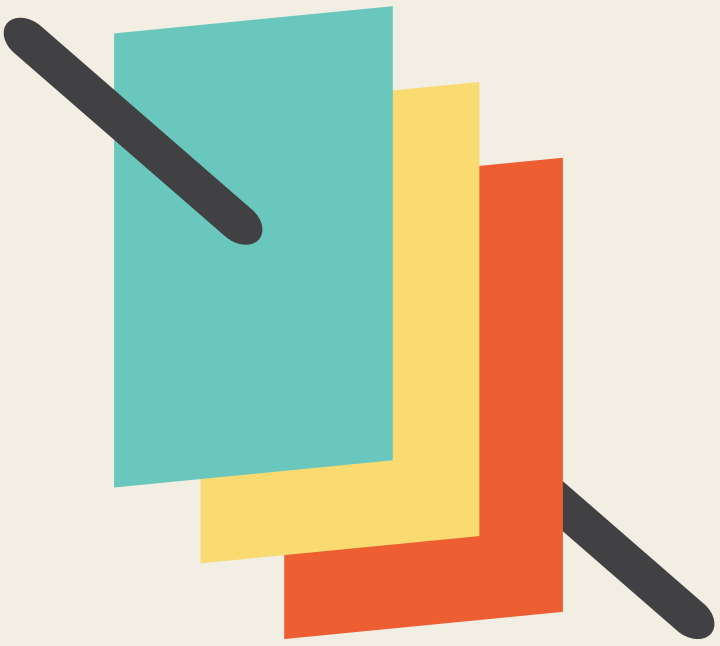
But this is just the beginning. The program is expected to remain a work in progress, always adapting to the community's requirements.

"To that end, we plan on sending annual assessments to the Texas data community to make sure that we're staying abreast of the community's evolving needs and their priorities," Cooke said.

Data captured from the pilot phase, which included participants from 28 agencies and institutions of higher education, indicate that the program will be a success.

"The high completion rate and positive reviews from everybody are compelling proof that it's effectively promoting data competency and providing immense value to the learners," Cooke said.

"By equipping participants with these essential data skills, the program's helping to empower them to make more informed decisions, improve work efficiency and enhance their overall professional development," he said.



How to Stop Worrying About IT and Focus on the Data



An interview with Matthew Alexander, Field Solutions Architect, Pure Storage

Legacy storage technology can be such a drag. It can be a drag on the performance of your overall IT environment, and on the efficiency and sustainability of that environment.

Matthew Alexander, now with Pure Storage, learned that firsthand when he was a systems administrator/ architect with a government agency.

The agency shifted to flash-based systems from Pure Storage. Flash storage has no moving parts, so it is faster and requires less energy. Flash also is much denser, storing the same volume of data in a smaller system. All told, a flash-based system consumes 90% to 97% less electricity to store the same amount of bytes.

For Alexander's agency, it was a revelation.

"It's interesting that when we incorporated flash storage in our environment, we didn't realize how much our previous storage platform was negatively impacting the performance and the availability of our environment, and ... how much operational complexity we had," Alexander said.

The Value of Simplicity

Complexity is a common problem for agencies with older technology. In many cases, for example, they manage multiple storage systems, each for a different type of storage format — e.g., block, virtual machine, file, or object. Each format has benefits, but it can be a bear to manage all those individual systems and the interfaces involved.

That's why Pure Storage offers an integrated data platform with an operating system called Purity that manages that complexity.

Another challenge is that agencies have integrated their compute, network and storage systems. That architecture requires an agency to scale up all three technology stacks in lock step. But that can be a waste of money. For example, you might need more storage but can work with the existing compute and networking capacity.

"Being able to scale granularly where you need it can help with budgets," Alexander said.

The Weight of Technical Debt

Another common problem is what's known as technical debt. Rather than investing money in new technology, an agency keeps investing in the legacy tech, scraping by as long as they can.

Agencies need a storage solution that evolves as requirements evolve, Alexander said.

"It needs to be able to support the agency's needs today, and fit within their current technology architecture, but not hinder their ability to move to the architecture that they will need in the future," he said.

That's the idea behind Pure Storage's Evergreen program. Through this program, an agency can purchase a storage platform today, and Pure Storage will keep the hardware up to date over time and add new software features as they become available. And unlike when buying a whole new system, there's no need to spend money on data migrations.

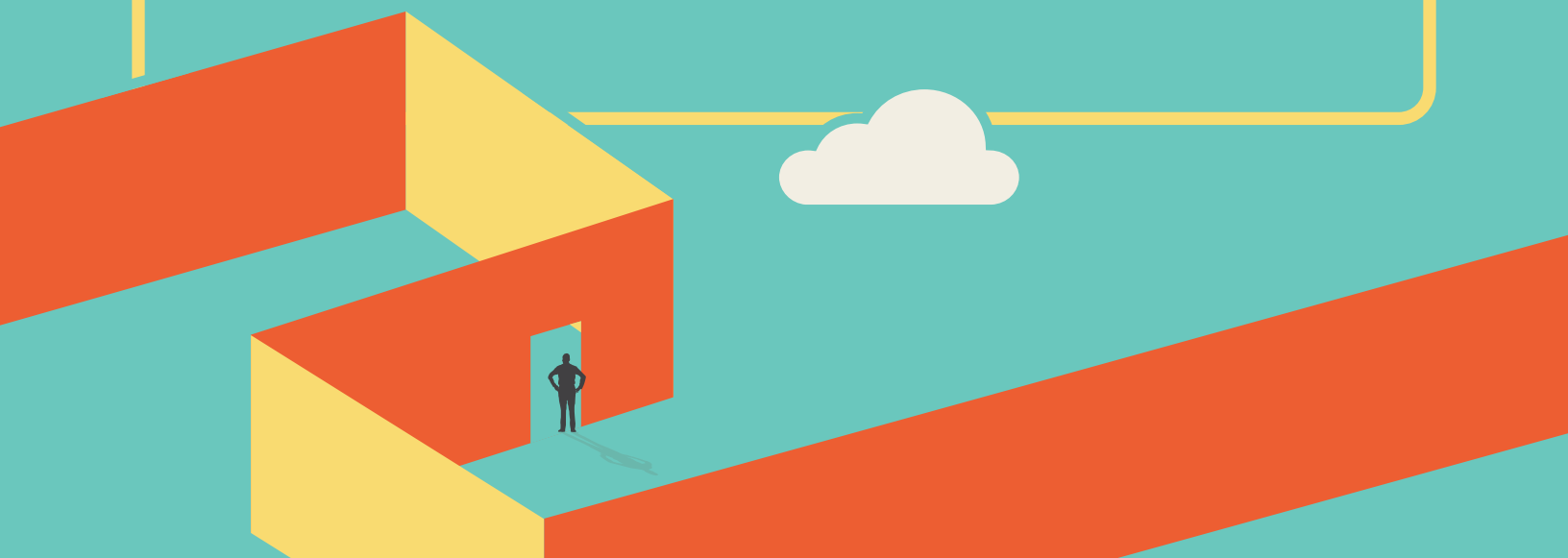
In short, by simplifying the IT environment and reducing technical debt, an agency can focus on "getting value from the data rather than managing the hardware on top of which it runs," Alexander said.

12 Defining Real-Time Data

Most data we rely on is static. It captures a moment in time — much like a photograph does — and can be outdated within weeks, days or even hours. That's OK in many circumstances, but it's a great disadvantage in others.

Real-time data, however, uses technology such as machine learning to compile, analyze and update data on a given subject. Sometimes the updates are near-immediate; sometimes they're simply more often than they used to be. Visual elements, such as maps and charts, in real-time dashboards help tell the story.

The goal is to better monitor constituent and other behaviors to develop more effective public policy.





In Fairfax County, 911 Dispatchers Track Real-time Data

One of government's core responsibilities is to help people in need. That's why officials in Fairfax County, Virginia found a way to ensure that local emergency response never wavers.

It's called RescueVision, and it's an application that gives the Fire and Rescue Department's 911 dispatch center real-time data on what's happening where. The program relies on geospatial information systems (GIS) technology, which identifies relationships between and among computer-readable datasets and produces a comprehensive, interwoven picture. In other words, GIS offers complex insight that humans, by themselves, cannot.

The RescueVision dashboard offers contemporaneous data on six areas of 911 concern:

Service coverage: How the county has distributed its emergency units

Resource availability: How many of each resource type are available to respond

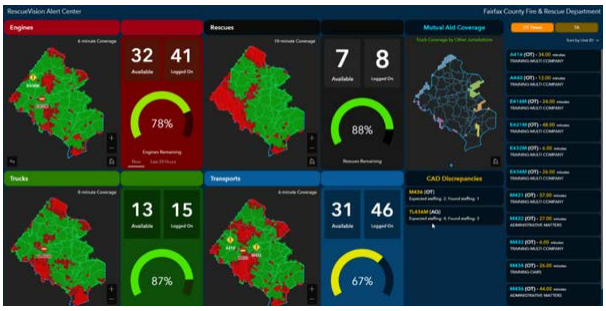
Units out of service: What resources are unavailable for dispatch, how long they've been that way and why

Transport location: How many emergency units are at medical facilities after transporting patients

Mutual aid coverage: What areas of the county other jurisdictions are serving

Discrepancies: What discrepancies exist between RescueVision and other fire and rescue systems

Multicolored maps provide coverage updates, graphs show key percentages (i.e., the percentage of response units in use) and a log tracks location and time details for each emergency unit — much like people track packages they order online.



Source: Fairfax County Government NewsCenter

The new system has notably improved the county's ability to serve its residents and demonstrates the government's belief in data-informed decision-making, county officials said.

"RescueVision shows how agile software development can come together with real-time data to provide actionable information right when it is needed," said John Morrison, the county's Data Analytics Strategy Manager. "We strive to be on the cutting edge of GIS technologies, and this is one of our many tools we use to ensure we're providing the highest-quality coverage to our residents and visitors."

Los Angeles Hears From Homeless Residents... by Cellphone Survey

Debates about homelessness often lack one key voice: that of unsheltered residents themselves. But in Los Angeles, which struggles to manage its many homeless encampments, officials devised an innovative way to hear from those folks — and in more real time than previously possible and regarding regulations and other concerns that call for quick action.

The city sends mobile phone surveys directly to its homeless constituents.

Research indicates that 94% of the city's unsheltered residents own cellphones — a surprisingly high number that's due, in part, to inexpensive options and government programs that provide free devices.

In the initial survey, a team from the University of Southern California (USC) and the University of California, Los Angeles asked homeless individuals about their health, living conditions and housing preferences. Results showed that the city's homeless are five times more likely to be food-insecure than the general L.A. population, three times more likely to be in poor health, and two times more likely to be anxious and depressed.

"The researchers were able to provide new information that just simply hadn't been provided, but also help us learn that it is possible to collect data in a less costly way than trying to go out to the streets and interview people," said Gary Painter, Director of USC's Homelessness Policy Research Institute, which released the initial findings.

Ninety percent of respondents want permanent or temporary housing, and only 2% would consider group shelters. The overwhelming aversion to group settings is largely due to safety concerns, lack of privacy and negative interactions with staff.

"There's this sense that as long as you get people under a roof, that's all that matters. It doesn't matter the quality of their experience under the roof. Just get them under a roof, and they can figure it out themselves," one survey respondent wrote.

Other surveys are planned throughout the year — "real time" means periodic, rather than annual, as it did in the past — and Los Angeles will measure the evolution of people's feelings and circumstances. Questions also will explore emerging issues, including the city's controversial encampment ordinances that restrict where homeless people can stay.

4 Tips for Using Real-Time Data

- **Identify issues** for which static data is too old to be truly useful. Where are your mission priorities and data resources misaligned?
- **Consider the data** you want to use and decide what "real time" means. Do you want constantly updated information, such as minute-by-minute stock market fluctuations? Or is fresh data every month OK?
- **Work with a vendor** to create a technological approach that collects all the information you want when you want it and presents the data in ways you prefer. Real-time analytics technology can be complicated, so find a vendor that can speak in a tech language you understand.
- **Train employees** to use and appreciate the real-time system. Switching from a single-instance data source to one that updates frequently can require a new way of working.



Finding Data Relationships

An interview with Rock Regan, Director of Alliances, State and Local Government, Qlik



Incomplete data can have dire consequences. It can prevent a child welfare caseworker from having an accurate understanding of an at-risk child's environment. It can leave a police officer unprepared for what waits on the opposite side of a door. And partial data can deprive needy families from services that would put food on their table.

Government is complex and inherently siloed, said Rock Regan with Qlik, which offers data integration and analytics. There are also many different agencies, each addressing a unique concern.

"The number one challenge that most governments face is they all want to be data-driven, but the data tends to be siloed and difficult to access. That makes agencies very stovepiped with the data they are able to use to make decisions," he explained.

Use Hard-to-Find Data

Legacy systems are part of the problem, Regan said, because they weren't designed for today's exponential data growth. What agencies need is technology that can access difficult-to-find information and present it to employees in ways they understand.

"Data is all about relationships," he noted, "so being able to figure out how to leverage data from multiple different sources and create data models that can help people make decisions they've never been able to do before is exciting."

Qlik, for instance, offers agencies what's called an associative approach, which lets them freely explore all their data. Qlik's platform is available in a FedRAMP moderate cloud and on premises, providing not only flexibility, but also the security, governance and innovation agencies require.

Employees, regardless of their technical knowledge, can use an interactive dashboard to explore data without limits, rather than rely on more finite information gleaned from specific searches.

That means employees can acquire ancillary knowledge they didn't realize they needed. It's like having peripheral vision.

Provide Strong Services

Regan said the objective is "to give knowledge workers in agencies the tools that allow them to be creative and to look at ways to provide better services to the public." That could involve public protection, benefit delivery or simply cost-cutting, he noted. Or it could involve real-time information sharing, such as when citizens want the latest election statistics.

One of Qlik's partners, for instance, helped the state of Georgia build what they call an online election hub, Regan said. As an election unfolds, people can check the Secretary of State's website for vote totals, number of absentee ballots cast, voter ages, and similar details. In a state such as Georgia, that information is of great interest.

That's an external-facing application. Behind the scenes, separate Qlik applications help Georgia counties perform their election responsibilities, Regan said.

Think About What You Want

As government, technology and data become more complex, agencies need to ease employee workloads, drive better insights and maximize resources. They must understand privacy and data-sharing rules.

And Regan said that agencies should dream big: "It starts with identifying what data they need. What does perfection look like?"

16

Defining Open Data

The term refers to data that's freely distributed and available for reuse and redistribution. It's often collected and compiled through contributions from multiple stakeholders or researchers, plus community members. Open data promotes collaboration and accelerates discovery — and helps agency directors make decisions based on the latest information.

The skillful use of open data repositories can encourage equity, improve public health, enhance the natural environment, stabilize communities, increase collaboration, fuel the economy, and advance research, public programs and quality of life.

All federally funded agencies with research programs now must release data to the public, and there is new guidance to help with managing and sharing data equitably. In addition, directives from the White House and key agencies have emphasized open data principles, and abiding by them helps build trust with the public and among other data users.



CDC Data Maps Illustrate Threat Impacts

It's often impossible to confine environmental and public health events to a specific jurisdiction, agency or area of responsibility. That is particularly true of disasters with wide-ranging impacts.

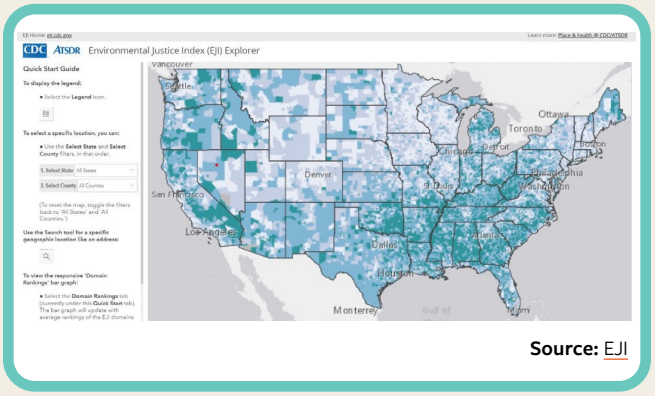
Last summer, the Centers for Disease Control and Prevention (CDC) launched its Environmental Justice Index (EJI), a new mapping tool that uses 36 indicators to measure environmental, social and health impacts nationwide. It is the first national geographical tool that measures — and releases to the public — cumulative environmental impacts and related public health vulnerability.

EJI draws its data from the CDC, Census Bureau, Environmental Protection Agency, and Mine Safety and Health Administration. The index was developed in collaboration with the Agency for Toxic Substances and Disease Registry, and it ranks communities based on pollution levels, environmental hazards, poverty, quality of education, walkability and other factors.

"Addressing environmental injustice is critical to advancing health equity," said CDC Director Rochelle Walensky in a press release announcing the index. "CDC is taking action to address the adverse health effects associated with environmental injustice by identifying those most at risk with tools like the Environmental Justice Index."

In a similar vein, several multi-agency initiatives have produced open data maps. One is the National Interagency Fire Center ArcGIS Online Organization, which pools mapping and evacuation data to help guide wildfire responses among state, local, tribal and national agencies. Another example is the National Integrated Heat Health Information System's Heat.gov, which includes a variety of temperature maps, including the National Oceanic and Atmospheric Administration's (NOAA) Urban Heat Island Map.

EJI and similar resources can help agencies with disaster response and guide them during normal operations when they're designing future projects, evaluating risks or speaking with a community.



Source: EJI

Texas, Feds Fight Opioid Crisis With Open Data

Last year 81,000 Americans died of opioid overdoses, and more than twice as many suffered a nonfatal overdose — and those were only the recorded cases. Open data resources that include community reporting of otherwise unreported events could be the key to reducing such tragic numbers.

According to Dr. Rahul Gupta, Director of the Office of National Drug Control Policy, the number of nonfatal overdoses today could be one of the most accurate indicators of fatal overdoses in the future. But until recently, most nonfatal events have gone undocumented — often because victims did not visit a medical facility that would record the incident.

Local research groups and community organizations have initiated programs to fill that information gap. TxCOPE, an open data app developed at the University of Texas in Austin, is one of them. It stands for Texans Connecting Overdose Prevention Efforts and allows community members to report overdoses anonymously, while aggregating traditional data sources. Because it will include a real-time map about where overdoses are occurring, the app helps direct emergency support to the appropriate areas, and it keeps agencies that collect overdose data equally informed, which is especially important in counties that provide scant access to toxicology reports.

"Data drives policy, data drives action," said Kasey Claborn, TxCOPE's Principal Investigator, in a Spectrum News profile of the project. She said the app empowers the local community to help address the opioid epidemic.

In December 2022, the White House launched a similar nationwide project, the Nonfatal Opioid Overdose Dashboard. Using information provided by emergency medical service (EMS) workers, the dashboard shows the number of nonfatal opioid overdoses in states and counties, the average number of naloxone doses administered and the average time it takes EMS to reach a patient.

Gupta said this dashboard will give local communities and agencies new ability to build their overdose response plans and evaluate their progress with reducing overdoses.

4 Tips for Using Open Data

- **Evaluate your repository.** Data is useful only if it's accessible and well-maintained. Evaluate your repository using the guidelines developed by the White House Office of Science and Technology Policy.
- **Educate agency staff, the public, stakeholders, elected officials and businesses about your data.** Data-driven decisions are the goal in your agency and for other agencies, elected officials, individuals, communities, entrepreneurs and businesses. Don't keep your valuable work a secret!
- **Provide storytelling or context.** Sometimes the collected data isn't enough to be meaningful to other agencies and the public. Make it count by providing context about why and how it was collected, and what questions it addresses.
- **Focus on trust.** Communicate clearly with your agency partners and community about the findings that may benefit them and their work. Interview users routinely to confirm that the data is accessible and understandable.

Analytics Draw a Complete Data Picture

An interview with Andy MacIsaac, Solutions Marketing Director for the Public Sector, Alteryx



alteryx | The Thrill of Solving

Spreadsheets are structured things: They have clearly defined lines, cleanly labelled columns, and rules that govern what goes where. For decades, government analytic programs have become skilled working within those parameters, even if it means spending countless hours manually manipulating data to fit. Spreadsheets are what Andy MacIsaac with Alteryx calls "30-year-old desktop technology."

But other data exists, doesn't it? The world is full of PDF documents, audio and video files, social media posts and other "messy" data sources, called unstructured data, that most agencies overlook. And most agency analytics programs are fragmented and overly manual.

Making Connections

"We need to modernize ... and democratize the approach to analytics within organizations," MacIsaac said, to eliminate the artificial divide between data experts and the people on the business end of agency operations. Organizations must curtail their use of manual systems and leverage all available data, structured and unstructured, as well.

"It's all about using that data to create ... the complete insights that you need to make sure people are getting the support and services they need ... and uncover potential problems that maybe you never thought existed," he explained.

Alteryx — which MacIsaac heard one client describe as the "Swiss Army knife" of data analytics — ingests data from different sources and manages it throughout an analytics lifecycle. That includes cleaning "dirty" data (e.g., removing duplicates) and applying descriptive, predictive, geospatial or other analytic approaches, using a single platform, he said.

A Real Difference

Consider the state of Ohio. Residents often were unaware that they qualified for more than one social services program, such as Medicaid, the Supplemental Nutrition Assistance Program (SNAP), the Temporary Assistance for Needy Families (TANF) and the Women Infant and Children (WIC) program.

Alteryx, through the work of a solution provider, helped Ohio compile enrollment data from multiple programs, matching identities, eliminating minor discrepancies and cross-checking information. The state identified and contacted roughly 77,000 people who likely qualified for benefits they were not receiving.

MacIsaac said the overriding question was, "how do we ... increase the social safety net, to make sure that the most vulnerable of our population are getting the services they need?"

Empowering Employees

The public sector has trouble attracting and retaining top talent, so managers should reevaluate how much time their existing employees spend on manual, mundane tasks that pull them away from higher-value work, said MacIsaac. And with a shortage of data scientists, agencies should adopt analytics platforms that everyone can use.

"You have some very capable people within your organizations," he observed, "[and] if you ... enable them with the right technology, they're going to be able to do some pretty impactful things."

Regardless of technical acumen, an individual can use an Alteryx analytics platform to answer questions that drive mission outcomes. In turn, that enhances employee skillsets, democratizes agency data, and makes analytics a team sport, he said.

Data is not theoretical, MacIsaac explained. Its impacts are real.

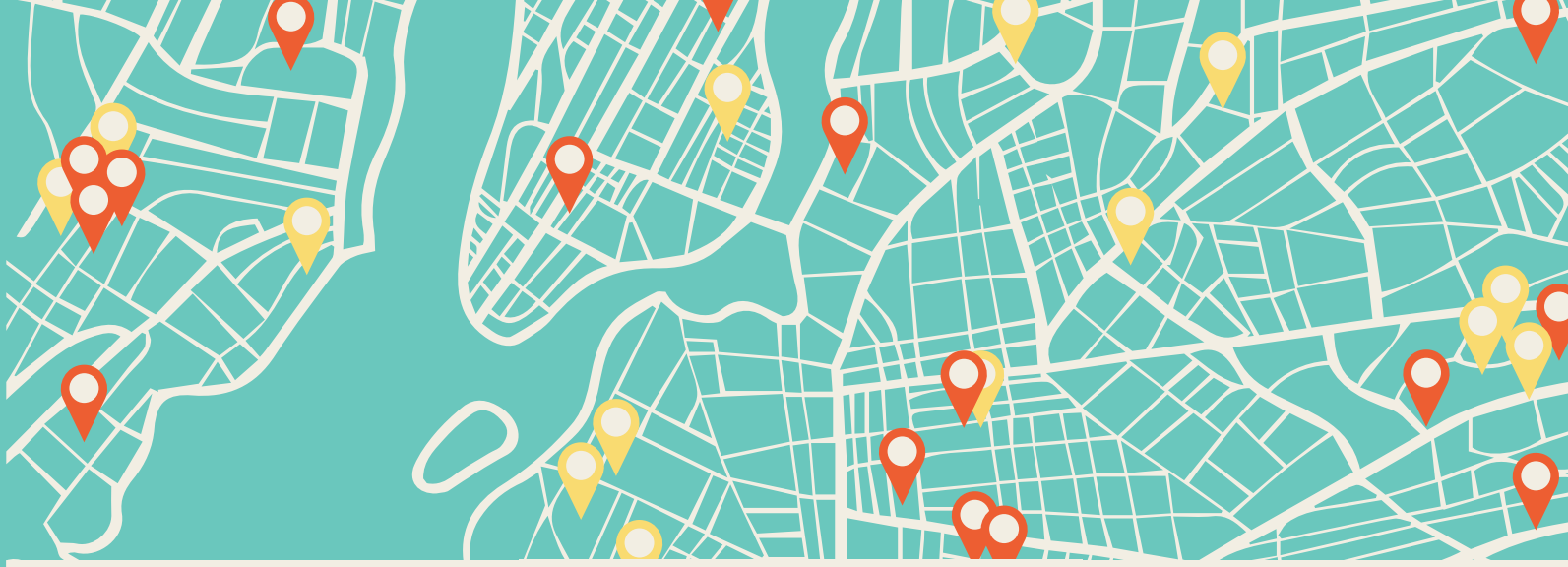
Defining Data Equity

20

The term "equity" is practically a buzzword these days — and it's a good thing to have top of mind. Synonymous with fairness, equity encompasses more than that. It's about the "consistent and systematic fair, just, and impartial treatment of all individuals." And when applied to data, it can "illuminate opportunities for targeted actions that will result in demonstrably improved outcomes for underserved communities."

Since President Biden signed the "Executive Order on Advancing Racial Equity and Support for Underserved Communities Through the Federal Government" in 2021, more than 90 federal agencies have created equity action plans. State and local leaders are acting, too. Here are three examples of data equity in action.



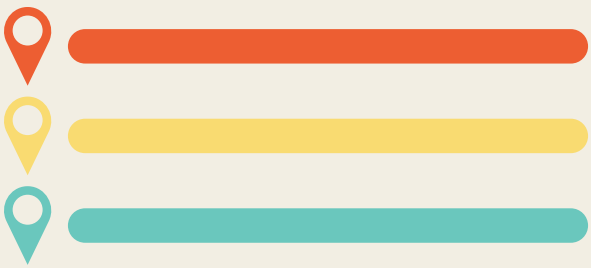


Food Pantry Mapping in South Carolina

The state's Department of Health and Environmental Control (DHEC) and Clemson University's College of Behavioral, Social and Health Sciences launched in 2022 an interactive ArcGIS-powered map that shows where people in need can get assistance. The EJ Strong Food Access Map — named after DHEC's Environmental Justice (EJ) Strong training initiative in underserved counties — covers all the state's 46 counties and has pinned more than 900 locations offering help, such as food and clothing, according to an announcement from Clemson.

Clicking a pin on the map makes a window pop up with the name, address and phone number of places offering help. Users can filter results by clicking an icon that will show what's nearby. They also can see whether a site has income restrictions or requires a photo ID or application, and get directions from the map.

DHEC notes that the map is extensive but not exhaustive. "It will remain a work in progress as more sites are added and the resource is kept up to date," the website states. The university maintains the map.



County-Level Community Exploration Through the Census Bureau

In April 2022, Census released My Community Explorer (MCE), an interactive dashboard with profiles of all 3,000-plus counties in the United States that shows location and demographic data on poverty, education access and language. MCE is designed to help users identify underserved communities and relies on four main datasets:

- Select demographic and socioeconomic statistics from the 2015-19 American Community Survey (ACS), which helps determine how $675 billion in federal and state funds get distributed each year.
- Risk factor group data from Community Resilience Estimates (CRE), which use 2019 ACS microdata and population estimates to measure individuals' and households' ability to handle the external impacts of a disaster.
- County Business Patterns, which includes data on the number of establishments, average annual payroll per employee and employment per county.
- Selected stats on self-employed workers from 2018 Nonemployer Statistics, such as the number of firms and average annual receipts per firm.

MCE users can explore by zooming and panning the map and clicking on a state, county or census tract. Clicking on one of those geographies brings up a window with 2019 CRE data, including percentages of the population with zero, one, two, or three or more predominant risk factors.

CDC Prioritizes Health Equity

Data practices that give everyone a chance at health became a priority when a CDC reorganization went into effect on Feb. 8, 2023.

"Our goal is an 'equity-centered' data system that accounts for social factors that have an impact on health, such as where people live, their environment, their income and jobs, the discrimination they face, and their access to health care," CDC's Walensky wrote in an email to the Atlantic Journal Constitution.

The shift follows CDC's Moving Forward effort in 2022 that included the creation of a new equity office to promote and monitor equity efforts departmentwide.

In recent years, CDC has faced criticism for its handling of the COVID-19 pandemic, which many say highlighted inequities in health care for people of color, those in rural communities and the unhoused — some of the factors in CDC's social determinants of health.

"We now have the opportunity as the science has increased, as we understand more about the drivers of health disparities, that we can also use our science to be more intentional about addressing those things," Leandris Liburd, Acting Director of CDC's new Office of Health Equity, told the newspaper.

Social Determinants of Health

Social determinants of health (SDOH) are the conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks.

SDOH can be grouped into 5 domains:

- **Education Access and Quality**
- **Health Care Access and Quality**
- **Neighborhood and Built Environment**
- **Social and Community Context**
- **Economic Stability**

Source: Health.gov

4 Tips for Ensuring Data Equity

- **Use existing datasets** from nonprofits, chambers of commerce and other community organizations to identify demographics that have a harder time accessing services and programming.
- **Provide forms** — digital and paper-based because not everyone has access to a computer — for collecting data from individuals in underserved populations in multiple languages, not just English.
- **Use disaggregated data** — data broken into subcategories — while ensuring the protection of private and sensitive information. Federal agencies can get assistance with disaggregating data from 13 statistical agencies. What's more, "statistical agencies should explore creating multi-year datasets for national surveys that will allow publishing estimates for small populations," according to "A Vision for Equitable Data."
- **Understand that equity efforts aren't one-time deals.** To ensure that data is accurate and actionable, collect and analyze it regularly, and engage community members, advisory boards, task forces and others in ways to take that data from statistics to actions.

Secure Data, Secure Decision-Making

An interview with Kevin Steeprow, Field Chief Technology Officer, Red River



Red River

Easy access to data is crucial for decision-making, especially in finding the right questions to ask. But security is equally crucial. As data increases, it becomes more challenging to store it, to share it and to secure it, within and across organizations.

"We want to use data more and more, to make data-based or data-oriented decisions," said Kevin Steeprow, Field CTO with Red River, a technology transformation company serving government and enterprise customers. "But we've still got that same dilemma that has existed since the dawn of time, that there's so much data out there. Where does it reside? Who has access to it?"

Red River helps agencies with many facets of technology integration that pertain to data, such as cloud migration, data center development, data mapping and infrastructure.

Take Stock of What You Have

To make solid steps forward, Steeprow recommended agencies "start where they are."

That means evaluating the systems that are already in place and finding out who can access them. Red River helps identify the characteristics of the highest value data sources to determine if they align with the agency's mission. Data that pertains to compliance requirements and mandates is a priority.

Visualization, which pulls together information from different sources, helps users understand what data means and respond to it.

Keep Access Authorized

A big challenge is managing the tension between keeping data secure and making it accessible.

Red River helps agencies build security models that help them with threat-hunting, through adaptive thresholding, mathematical principles and machine learning (ML), as well as dashboards highlighting potential malicious activity.

"So, with all those things to protect the data, how are we going to get access to it in order to make a decision that's meaningful and appropriate?" asked Steeprow.

For example, if zero-trust architecture prevents an agency from accessing data needed for decision-making, Red River systems help notify the user that data is missing, because of the way they have tried to access it. Steeprow and his team also help agencies record data correlations, so that there is a historical record of how users have analyzed information, he said.

Build a More Complete Picture

According to Steeprow, agencies need to get a full, holistic view of data systems, "rather than just diving in and asking a question, pulling data and coming up with an answer — and hoping that that's the right answer."

If someone is working with incomplete data, they might be asking the wrong question, resulting in useless, or worse, misleading results, he said. That's especially true for emergency situations when directors need to make decisions quickly.

As Steeprow said, AI and ML are only just beginning to bring data management to a place where information is comprehensive and accurate. But already the process is helping software engineers and agencies think in new ways, ask questions and evaluate the processes.

All those capabilities — when fully integrated into an agency's data coordination and processes — lead to more secure and complete use of data.

Finding Data Catalogs Drive Agency Missions

An interview with Scott Woestman, Vice President, Sales, U.S. Public Sector, Alation



Throughout the day, each of us makes decisions based on available data. And while some of those choices have little consequence, others are significant. Think about all the precise, real-time data the U.S. needs to logistically support today's military, for instance, or to respond quickly after a natural disaster.

"The challenging thing about data [is] it grows exponentially," said Scott Woestman of Alation, a data intelligence firm. "However, more data provides more opportunities for agencies and programs to solve mission-critical problems that impact all of us as citizens."

Agencies deprived of strong, plentiful information can struggle, he said, with employee hiring and retention, logistical support of core initiatives, and improper financial payments.

In fact, according to one estimate per the GAO, the federal government made nearly $281 billion in erroneous payments in FY21 that it cannot retrieve. That's a "remarkable amount that's not even due to fraud," Woestman observed. "More often than not, it's due to bad data." High-pressure working conditions only compound the problem.

"Government employees, just like a lot of corporate employees, are overworked," he said. "They've got to make quick decisions, and they don't often have the best actionable data to work with, and mistakes get made."

Catalog It

Now, imagine if employees had a solution to support data-driven decision making which combined the features of Amazon.com with the Dewey Decimal system. That's how Woestman describes a data catalog. Using machine learning, it creates and maintains an enterprise-wide inventory of trusted data that people across an agency can search, access and use to collaborate, rather than relying on decentralized data sources.

Gartner calls data catalogs "the new black in data management and analytics" and says they are "fast becoming a must-have tool." According to Woestman, they strengthen operational intelligence so government can respond when, where, and how it needs to.

In Practice

For instance, during the pandemic, Alation partnered with a federal healthcare agency to develop an enterprise data catalog that facilitated sharing information across their siloed centers. This enabled the agency to help save lives during the pandemic and, in the future, use key data for other lifesaving programs. The agency had tons of data, but it needed Alation's guidance to identify the most actionable, useful information, Woestman said.

Employees looking for a catalog solution have a simple case to make for adopting one, he said: Data catalogs will help them work better and faster and make their agency more effective.

Governing Well

People often are less enthusiastic about another key element of data strategy: governance. Just ask chief data officers who broach the subject with their employees, Woestman said. To make governance part of the workforce culture, he suggests getting employee buy-in early on.

"Data governance doesn't have to be punitive," he offered. "We can empower [people], and we can help them collaborate and drive results to build their own governance paths.... You've got to start with the employees."

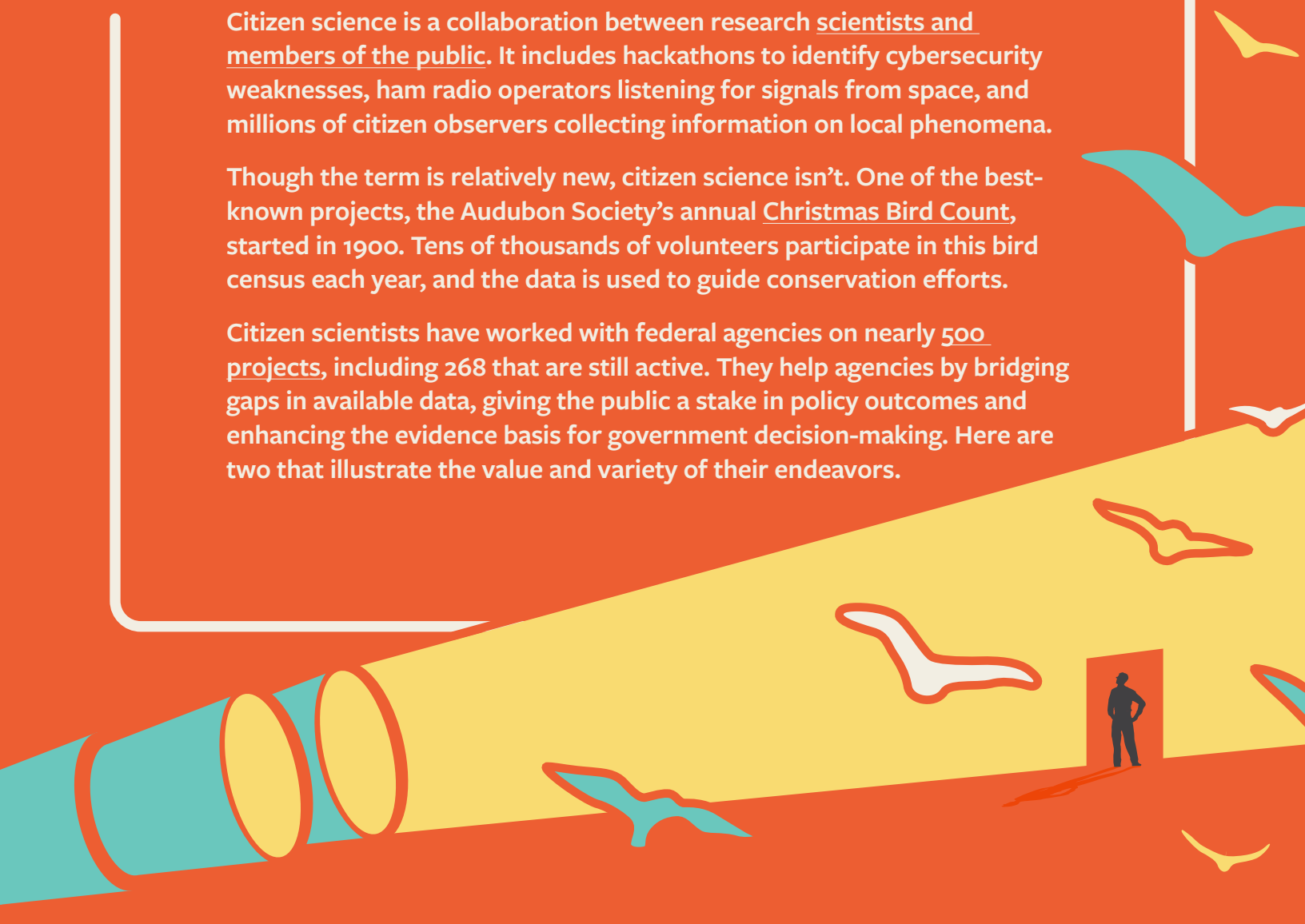
Defining Citizen Science

25

Citizen science is a collaboration between research scientists and members of the public. It includes hackathons to identify cybersecurity weaknesses, ham radio operators listening for signals from space, and millions of citizen observers collecting information on local phenomena.

Though the term is relatively new, citizen science isn't. One of the best-known projects, the Audubon Society's annual Christmas Bird Count, started in 1900. Tens of thousands of volunteers participate in this bird census each year, and the data is used to guide conservation efforts.

Citizen scientists have worked with federal agencies on nearly 500 projects, including 268 that are still active. They help agencies by bridging gaps in available data, giving the public a stake in policy outcomes and enhancing the evidence basis for government decision-making. Here are two that illustrate the value and variety of their endeavors.





Great Smoky Mountains Host Citizen Observers

The National Park Service (NPS) fields more than two dozen citizen science projects, both park-specific and multi-park. From Denali in Alaska to Saguaro in Arizona, volunteers collect data, upload it, and share in the scientific achievement.

At Great Smoky Mountains National Park (GSMNP), which spans the North Carolina/Tennessee border, volunteers record flora, fauna and fungi using an app that geotags their photos and attaches other information about where the photos were taken. The app then uploads the images and information to a database.

So far, volunteers have helped map over 10,790 of the 21,500 species known to inhabit the park, which is the most biodiverse of all NPS locations. They've discovered at least 10,790 previously not recorded at that location.

The citizen scientists' contributions are part of a larger project called the All Taxa Biodiversity Inventory, which began in 1998 and is coordinated by the nonprofit Discover Life in America (DLiA). The inventory is now the largest sustained natural history inventory in the United States and one of the largest in the world, according to the park's website.

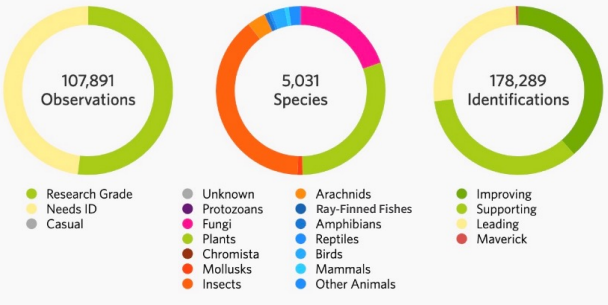
Park scientists use the mapped data to learn where species live, the ecosystems in which they play a role and environmental factors that may affect their populations — information that guides conservation efforts.

GSMNP volunteers are recruited on the park's website, the DLiA website and elsewhere online. The process is simple: People sign up at iNaturalist and download the app, which they can use to participate in GSMNP and in other tracking projects worldwide.

To keep citizen scientists involved, a dashboard shows photos of recent observations and a leaderboard of the most observations and most species. There's also a tally of frequently seen species. And every mention of the project emphasizes how important volunteers are.

"INaturalist usage in the Smokies has skyrocketed from just four users in 2011, to 3,800 in 2020, to now more than 7,100 users," said Will Kuhn, DLiA's Director of Science and Research, in an interview last September. "We need visitors' help making our diversity in iNaturalist to match our true diversity."

Contributions by iNaturalist Users to Great Smoky Mountains National Park All Taxa Biodiversity Inventory



Source: iNaturalist

Massachusetts Citizens Take Their Local Temperatures

Excessive heat causes more deaths in the United States than any other weather condition, so NOAA's National Integrated Heat Health Information System and its partner, CAPA Strategies, are organizing citizen volunteers to document urban heat islands — areas where temperatures can rise 10 to 20 degrees Fahrenheit higher than in surrounding neighborhoods.

More than 60 Massachusetts communities have mapped their hot spots. Cities use the data to develop heat action plans, add cooling stations, focus urban forestry efforts, and support policy, research and education.

As part of that program, 80 volunteer scientists measured temperatures, humidity and air quality levels along the Mystic River Watershed, which covers 76 square miles in 21 municipalities around Boston, in a project dubbed "Wicked Hot Mystic."

Boston's Museum of Science (MOS) partnered with community organizations, including the Resilient Mystic Collaborative, Mystic River Watershed Association (MyWRA) and Metropolitan Area Planning Council, using tools from CAPA Strategies. Officials began recruiting volunteers in January 2021 for the August collection. The campaign involved event outreach, emails to MOS and MyWRA members, social media posts, flyers, and public education.

On Aug. 12 and 13, 2021, volunteers drove or biked through specified locations at 6 a.m., 3 p.m. and 7 p.m., and then at 6 a.m. the next morning. They carried heat, humidity and particulate matter sensors and GPS units that collected data once every second along the route. Each participant received a $25 gift card.

The survey showed temperature differences of up to 10 degrees Fahrenheit. The collected data was used to create high-definition maps of each time period's results and mapped correlations between ground-level air temperature and both physical and social factors. These included impervious surfaces (i.e., roads and parking lots), tree cover, air quality, and previous redlining by mortgage lenders and government programs.

The data has been presented widely to local Massachusetts communities and will inform climate mitigation efforts throughout the region.

4 Tips for Encouraging Citizen Science

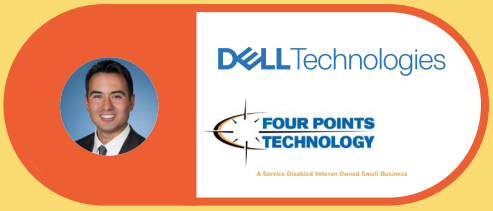
- **Choose a project that's significant to the community.** Citizen science depends on volunteers, so what's being measured should be something people can get excited about.
- **Partner.** Federal, state and local organizations all have roles in citizen science. Connect local projects with federal initiatives and connect federal projects with community groups on the ground. Nonprofits with technical expertise are also important players.
- **Take advantage of existing tools.** Phone apps like iNaturalist and iSeeChange are easy to use, and they link citizen data with national and international databases. The Citizen Science Alliance helps recruit volunteers and has tools for setting up projects.
- **Report and act.** People need to see that their work was relevant and that the results will be used in concrete ways. It's important to connect the data collection with real change.





Simplifying Backup in a Complex Data World

An interview with Brad Montgomery, Director, Federal Presales at Dell Technologies



Agencies are using data in new ways — sharing across departments, in the field, and with remote workers and the public. That unleashes the power of data to serve the mission, make informed decisions and make programs more effective.

It also exposes the data to more risks, and not all of them can be avoided. That means you need to plan not just for prevention, but for recovery.

Data is an agency's most critical asset, said Brad Montgomery of Dell Technologies. "And if your data isn't protected, you can't put it into production."

Distributed Data Makes Protection More Complex

While hackers, ransomware and insider attacks are attention-grabbing threats, data is also vulnerable to more mundane risks — accidental deletions, misconfigurations, file corruption, failed hardware. "We've had customers that lost data because of a sprinkler system malfunction," Montgomery said.

Reliable backup is essential to recovering data and restoring operations. However, as agencies modernize and migrate to the cloud, existing data protection may not provide as much security as before. Many modernization projects forget to include upgraded backup. New applications or platforms and their service agreements may change who's responsible. Too often, agencies read the fine print of their cloud provider contract only to discover that it still rests on them to protect their data, Montgomery said.

Where agencies use the data has grown harder to keep track of, too. "Data does not live just in the datacenter," Montgomery said. "It's in the cloud, it's in the remote offices, it's on mobile devices. And as that data becomes more distributed, it's harder to protect."

How Dell Technologies Helps

Dell Technologies offers a comprehensive product that also reduces complexity. APEX Backup Service is cloud-native and designed to protect software-as-a-service (SaaS) applications, on-premise and cloud-based workloads, and all endpoints, including laptops and mobile devices, according to Montgomery. It supports role-based access control and multifactor authentication, helping secure data wherever it's being used.

"We have anomaly detection to catch signs of intrusion, quarantine capabilities, and curated snapshots that let you roll back in time to the last known good copy," he said.

Since the software is delivered as a service, it requires no added hardware — speeding up implementation and relieving the complexity introduced by backup infrastructure. Most importantly, it brings backup protection into a single framework and administration into a single management console.

Dell Technologies' APEX also addresses increasing security concerns: It meets Federal Information Processing Standard 140-2 encryption requirements and is authorized under both Federal Risk and Authorization Management Program (FedRAMP) and the Association of International Certified Professional Accountants (AICPA) Systems and Organization Controls 2 (SOC 2).

And, APEX addresses the threat that's keeping agency leaders up at night: ransomware. "We have modern tools to fight ransomware," Montgomery added. "Everything is air-gapped and immutable. Metadata is separated from the data shards themselves. [It's] a holistic approach to solving those data protection challenges."

How to Innovate in Cybersecurity (and Protect Data Assets)

Insights from Technology Integration Group

Nowadays, it is no secret the public sector needs more cybersecurity innovation. From the top level down, agencies nationwide are endlessly defending their resources from cyberthreats — and cyberthreats are constantly reinventing themselves, so agencies must keep up.

The danger perhaps is greatest for agencies that take advantage of new opportunities to identify, capture and analyze data, including from less traditional sources, such as online audio and video files.

Yet innovation is difficult without resilience. Resilience is the ability to respond to, recover from and continuously function during disruptions; without resilience, agencies may find innovation hard to reach.

Agencies can acquire both attributes by imitating the private sector. Take Technology Integration Group (TIG), an IT solutions provider. Its "Start Right" philosophy is a business methodology that can make agencies more innovative and resilient.

Here are five ways agencies can become resilient innovators, according to TIG:

1. Engage with Challenges

Before agencies innovate, they must clearly understand their business needs. First, they must decide how and what innovations may elevate their credibility and work. Next, these agencies must determine which of their existing capabilities already contribute toward the potential innovations. Finally, these agencies should craft proposals that capture each innovation's impact.

Take data encryption, which converts data into a form that, ideally, only authorized parties can decipher. Agencies handle sensitive data about citizens, so protecting this information is vital for preserving public trust. Agencies without data encryption can start applying it by determining which information is most critical and may need encryption first.

2. Conduct Analyses Agencywide

Innovation also requires gathering as much information as possible about potential next steps.

Understanding their current operations, architecture life cycles and frameworks can not only assist agencies with refining their business needs but help them model their future routines.

For instance, look at data storage. Currently, many agencies have legacy IT that they cannot easily restore after cyberattacks. However, using cloud computing, agencies can store backup copies of their data and become more resilient.

3. Build Blueprints

All innovations, including the cybersecurity variety, need roadmaps. For the best results, agencies should measure how innovations might transform their operations. By predicting the potential benefits and financial impact of any changes, agencies can craft the best possible solutions for their workforces.

4. Make the Case

Innovators must articulate how the changes they are proposing can take root at agencies and what investments they need to enact them. Without these details, agencies may lack lasting innovation and resilience.

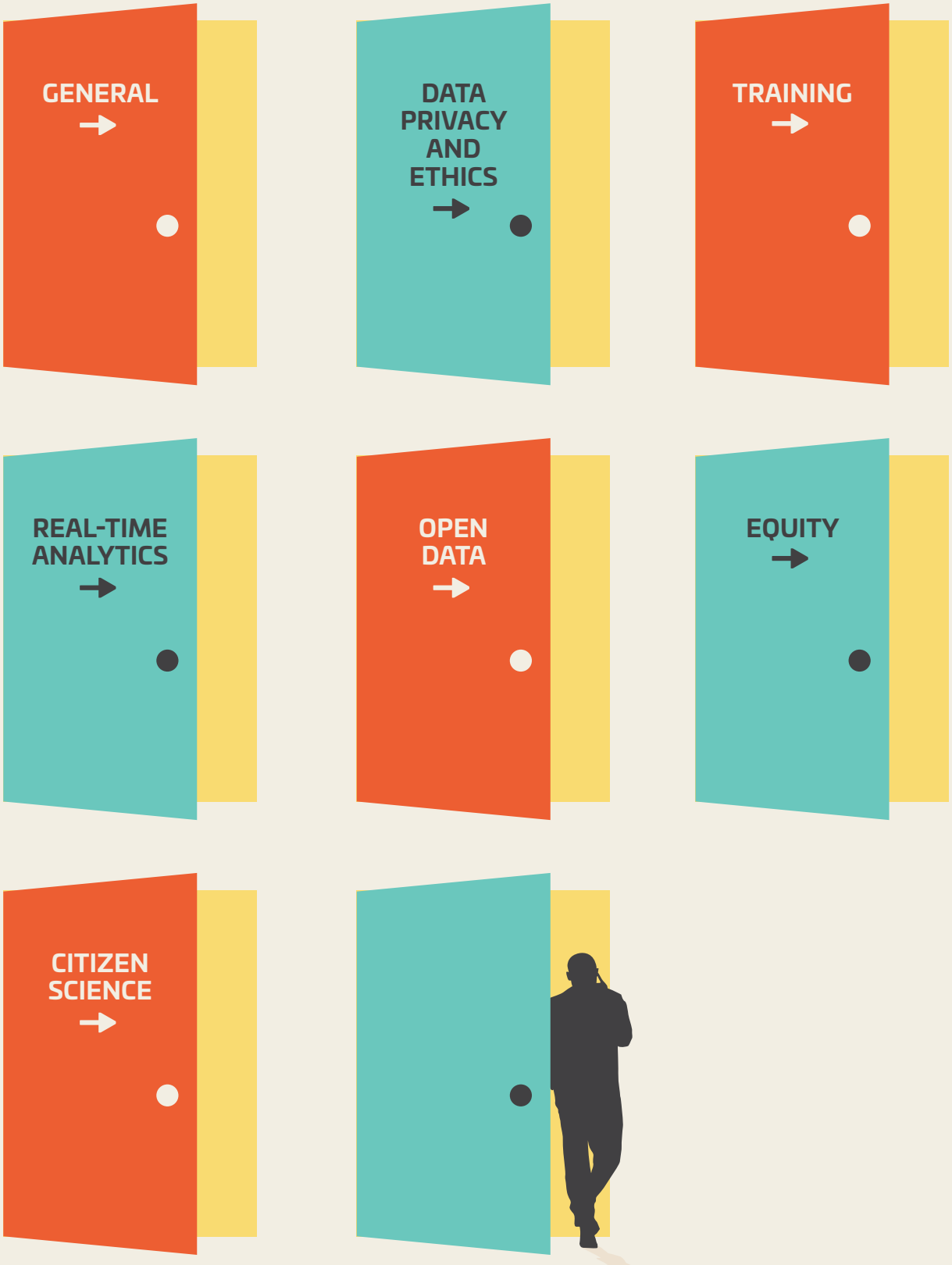
5. Implement Effectively

To optimize the benefits they reap from innovations, agencies must first determine the best way to implement them. Ideally, implementations unfold using roadmaps tailored to produce the maximum number of advantages from the solutions involved.

Innovation may seem difficult on paper, but partners like TIG can demonstrate what consistent, simple innovation that supports data collection and analysis looks like. Over time, innovative agencies have the agility to stay resilient when cyberthreats come calling.

Additional Resources

Want to dig deeper into the topics explored in this guide? Here are some resources to get you started.



Takeaways

Data offers agencies so many possibilities for better-informed decision-making. This guide showcased five opportunities.

Employee skills. Assess your staff's data literacy levels, then develop programs to fill the gaps, including online and individual agency learning programs. Strengthen employees' ability to read, understand, work with and communicate with data.

Real-time data. Identify where your mission priorities and data resources are misaligned, decide what "real time" data collection means for you, and work with a vendor to craft a system tailored to your needs.

Open data. Consider what freely distributed, well-maintained public data can further your agency's goals, encourage and appreciate public data collection that supports those objectives, and provide data context if needed.

Data equity. Use existing datasets to identify groups that struggle to access services and programming, and give underserved communities digital and paper forms, in multiple languages, to collect their information. Conduct ongoing data collection and analysis.

Citizen science. Choose a volunteer project of community significance and partner with federal, state, local and nonprofit entities as appropriate. Take advantage of existing apps and other tools, and let citizen scientists know the concrete results of their work.

Thank You

Thank you to Alation, Alteryx, Dell Technologies/ Four Points Technology, Pure Storage, Qlik, Red River and Technology Integration Group for their support of this resource for public sector professionals.

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.

Authors

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

Designer

Calista Lam, Graphic Designer





1152 15th St. NW Suite 800
Washington, DC 20005
P: (202) 407-7421 | Fw: (202) 407-7501
www.govloop.com
@GovLoop