YOUR DATA IN THE YEAR OF EVERYTHING ELSE

A GOVLOOP GUIDE

FAKE NEWS

STAY HOME
SAVE LIVES

VOTE!

HAVE YOU COMPLETED THE 2020 CENSUS
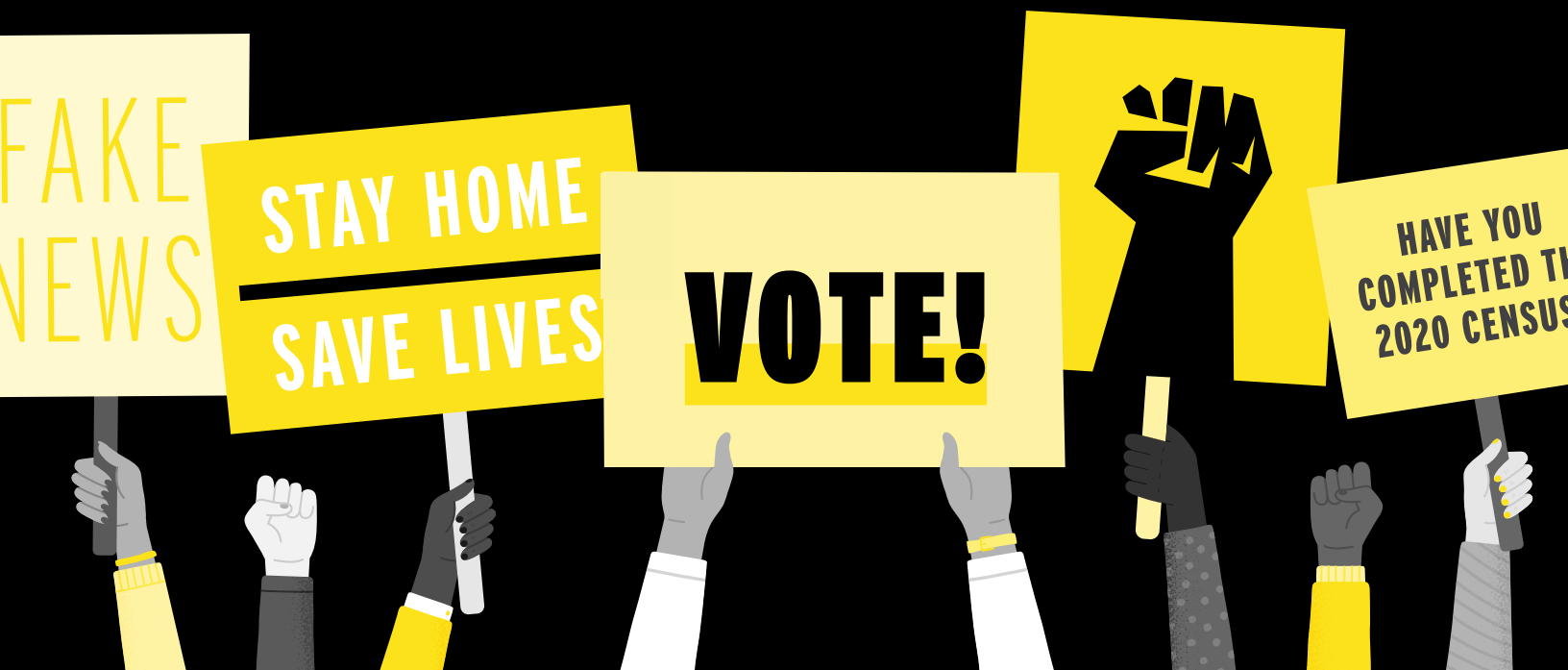
"GOVERNORS RARELY EVEN HAD USED THE WORD 'DATA' – MAYBE ONCE OR TWICE IF THEY EVEN SAID 'DATA' AT ALL IN THEIR STATE OF THE STATE SPEECHES. AND NOW HERE WE ARE. EVERY DAY, EVERY GOVERNOR IS USING DATA, RIGHT IN THEIR PRESS CONFERENCES, TALKING ABOUT DATA."

– Tyler Kleykamp, Director,
State Chief Data Officers Network

CONTENTS

EXECUTIVE SUMMARY

Late in 2019 and carrying into early 2020, speakers circulating government conference circuits pronounced 2020 "the year of data."

If only.

Nobody could have foreseen how torrid and trying a year it would be, but "the year of data" doesn't come close to encapsulating 2020.

Within months, the novel coronavirus had swept the globe, killing close to a million people – a heartbreaking number that is certain to rise much higher following the publication of this report. COVID-19's landfall on U.S. shores shuttered businesses and sent millions home without work. In response, Congress passed the Coronavirus Aid, Relief, and Economic Security (CARES) Act, which offered hardship payments and expanded unemployment benefits to citizens, while providing emergency funding to businesses and state and local governments.

In the midst of the pandemic, America also reckoned with the forces of racism and police brutality, brought to a head with the killing of George Floyd. Floyd, a 46-year-old Black man, was killed in Minneapolis when a white police officer knelt on his neck for eight minutes and 46 seconds while detaining him for allegedly using a counterfeit $20 bill. Floyd's death sparked a nationwide outcry and demand for reviewing police techniques, budgets and incidents.

Data has played and will continue to play a major part in understanding and responding to these pressing issues, which will remain relevant decades in the future. Though 2020 might no longer be "the year of data," data's importance and role in government has unquestionably swelled to meet modern moments, according to experts in the field. In 2020, data has informed the public, guided decision-makers and transformed government policies.

Employees who never had to think about data in their day-to-day roles suddenly find themselves assigned to studying the numbers. And policymakers who rarely mentioned data in addresses or agendas now have daily reports.

Consider how:

- Visualizing up-to-date and location-specific data, **coronavirus dashboards** are providing the public and decision-makers with simple tools for evaluating COVID-19 trends.

- Governments are assigning employees to new **data-involved roles**, and some are even creating offices focused on data.

- Data is being used to **analyze inequality** in the United States and determine where aid and reform is needed.

- **Partnerships** are popping up to meet data needs.

This report covers these trends, among many others. It features a variety of voices, from government experts to volunteers and advocacy groups, and provides relevant, timely tips for improving public sector data.

Though based on a moment in time, the scope of this report is in no way confined to 2020. The events of this year will live long in history books, and even if it's not "the year of data," government data will never be the same.

DATA IN THE NEWS

The landscape of government data is vast and highly individualized. These recent news clips, however, can help paint the picture of where government data efforts currently stand.

TESTS HIT SNAGS IN OLD TECHNOLOGY

Across the country, counties have been ill-prepared for the rapid ramping up of COVID-19 tests. As The New York Times reported, in Texas, test results have faced bottlenecks because of fax machines. Harris County, specifically, at any given time can receive hundreds of pages of test results via fax.

Travis County, Texas, meanwhile receives about 1,000 faxes a day, some of which are duplicate results. These faxed documents have quality problems and, while sent to a computer, still have to be printed out and manually coded into databases.

The problems with using fax machines are copious. Faxes can be directed to mobile numbers and wrong departments, or they can come with incorrect or incomplete information. For that reason, demographic data is often left off COVID-19 reporting, despite national efforts to study the impacts of the coronavirus on specific populations.

The lack of funding and technological resources is a major problem for local governments, many of which are facing especially meager economic circumstances due to the pandemic.

FEDERAL DATA STRATEGY ACTION ITEMS GO INTO PLAY, SOME DELAYED

In response to the coronavirus, the Federal Data Strategy team extended several deadlines for its action items. Agency action items have been delayed by up to a month, while one shared service action item was pushed back by three months.

Action items that were delayed include agencies' creation of data governance bodies and identifying opportunities to increase staff skills. Many requirement dates remain the same in the action plan's kickoff year.

The Federal Data Strategy was published in 2019 with the goal to "fully leverage the value of federal data for mission, service, and the public good." Other action items address conducting internal data maturity assessments and publishing open data plans. In addition to an action plan, the strategy also features principles and practices for federal agencies to adopt.

A welcome side effect of the Federal Data Strategy has been that more state and local governments are publishing their own data strategies. Several states, like Connecticut and Oregon, have alluded to the Federal Data Strategy as their blueprint.

FEDERAL DATA STRATEGY AGENCY ACTIONS

1.
Identify Data Needs to Answer Priority Agency Questions

2.
Constitute a Diverse Data Governance Body

3.
Assess Data and Related Infrastructure Maturity

4.
Identify Opportunities to Increase Staff Data Skills

5.
Identify Priority Data Assets for Agency Open Data Plans

6.
Publish and Update Data Inventories

QUESTIONS OF DATA TRANSPARENCY IN COVID-19 RESPONSES

Florida was the center of praise for its COVID-19 response, and event-goers began to flock to the Sunshine State's reopening party – that was, until a prominent Florida geographic information systems (GIS) manager accused the health department of undercounting and manipulating data to look more favorable. As The Washington Post reported, Rebekah Jones, who maintained a localized dashboard of COVID-19 test results and cases in her time as a GIS administrator, alleged that Florida officials asked her to manipulate positive test numbers and engage in dishonest data practices. The Florida Health Department fired Jones for "insubordination."

Among Jones' accusations, she alleges that health officials requested she manually delete cases, which would reflect fewer positive tests. She also said the state's number for tests was based on samples collected instead of individuals tested.

The White House had previously heralded Florida as a model of success for its data tracking and response to the coronavirus. But as the summer stretched on, Florida's cases surged to more than 15,000 in one day, setting a single-day record for any state in the country.

Nationally, the White House directed hospitals in July 2020 to circumvent the Centers for Disease Control and Prevention (CDC) and report information about patients and available resources to the Health and Human Services Department (HHS), of which the CDC is a component. The White House said the move would better inform its coronavirus task force and streamline the data collection process.

Many in the health information community, however, worried that the new protocols would lack transparency and take valuable information away from researchers and officials who rely on CDC data. Much of the original information HHS receives is not publicly available, and critics of the move worry about the politicization of data in a newly opaque environment.

The CDC has made missteps in tracking COVID-19 data as well, such as when it combined the information of patients with the coronavirus with those who had recovered.

2020 DATA TERMS

These terms may not be new, but in 2020, they rose to prominence in the data community.

Chief Data Officer (CDO)

A position generally responsible for data governance, coordination and prioritization that varies in scope across government.

Contact Tracing

The practice of identifying people who have infectious diseases and tracing their connections to minimize social contact and community spread.

Geographic Information Systems (GIS)

A computer technology that allows users to input, analyze and manipulate spatial and geographical data in conjunction with more traditional data forms.

Data Governance

The guidance, rules and policies applied to organizations' data to foster standardization, interoperability and usage.

Social Vulnerability

The resilience of a community when faced with exigent pressures.

FACT OR FICTION?

The coronavirus has disproportionately impacted some communities and demographics.

Fact: CDC data shows that Black and Latino people have contracted the coronavirus at significantly higher rates.

LIMITED RESOURCES. LIMITED TIME. MASSIVE AMOUNTS OF DATA. DISPARATE WORKFORCE.

Altair Data Analytics enables the public sector to arrive at actionable insights quickly.

▶ **Learn More – Altair Solutions for Government Organizations**

▶ **Welcome to No-Code, Automated Data Transformation - Start Your Free Trial Today**

△ **ALTAIR**

in. f ⅴ ◎ #ONLYFORWARD

LEVERAGING DISPARATE DATA IN A DISPARATE WORKFORCE

An interview with Harrison Murphy, Director of Data Analytics Solutions, Altair

Most agencies were built as extremely centralized organizations, where policy and direction trickle down from headquarters and data and results circle back around.

COVID-19 brought a shock to that system. Agencies converted to telework, and newly distributed structures saw data exclusively generated, processed and analyzed at the edge.

"The decentralization of agencies exacerbates the drawback inherent to the centralized system," said Harrison Murphy, Director of Data Analytics Solutions for Altair, a global data analytics software provider for government.

Agencies have no choice. They must adapt to an environment where employees can work with data from anywhere. When agencies fail to adapt, they risk bottlenecks in their processes.

In an interview with GovLoop, Murphy laid out three areas agencies should focus on.

1. PREPARATION

Data used to be generated through highly controlled organizational processes, such as logging or filing constituent information. Now, data creation is increasingly driven by individuals.

The keys to success in remote environments are flexibility and agility. Agencies need clear strategies and governance policies that dictate how data should be acquired, prepared and structured, so employees aren't left stranded without guidance. These groups also need systems to place controls on data while still giving employees the freedom to innovate.

2. COLLABORATION

Remote work realities mean employees can't peek over a cubicle and ask for help. Despite this barrier, collaboration and security are increasingly important as insights are locked away on individuals' computers. Agencies need a platform that allows users to work

on the same datasets at the same time, while also eliminating the security risks associated with housing data on employees' laptops.

"We see collaboration going beyond the simple passing of the data baton from one user to another," Murphy said.

In the office, a data analyst could share a multifunctional model on a hard drive. Those same capabilities are crucial remotely too, but they must be available in the cloud.

3. ANALYSIS

Data acquisition and preparation are critical components of self-service analysis, which unlocks invaluable information, gathers insights and identifies trends. In today's era of remote work, without the right tools, agencies' scales can tip even more in favor of preparation instead of analysis.

Non-tabular sources, or language-based data types such as online forms, often need to be coded by data scientists – unless agencies have a solution that removes the heavy lifting.

"Agencies need to find the balance between allowing end users agility and flexibility but still maintaining a governed, curated infrastructure," Murphy said.

With Altair, agencies' structured and multi-structured non-tabular data can be standardized and prepared for analysis in a user-friendly platform online. Then, data analysts can generate supervised and unsupervised models in the cloud – in a central location even though they're not in the office.

The final step unlocks visualizations and advanced analytics. But none of that's possible without the hard work of adapting processes to a distributed environment.

"Rather than trying to reestablish 'the way things used to be,' agencies may gain efficiencies by addressing the pain points of current data processes, allowing them to flex with future interruptions," Murphy said.

THE RISE OF DATA DASHBOARDS

Dashboards: An online interface that shows stats, graphics and visuals that are usually easy to understand and interpret for a broad audience.

State, local and federal agencies have deployed dashboards in response to the COVID-19 pandemic. New to some and not to others, dashboards present important data to guide decisions. Dashboards can be both internal and external. Here are some examples.

Images of these dashboards do not represent the most recent data. Go to the websites for updated information.
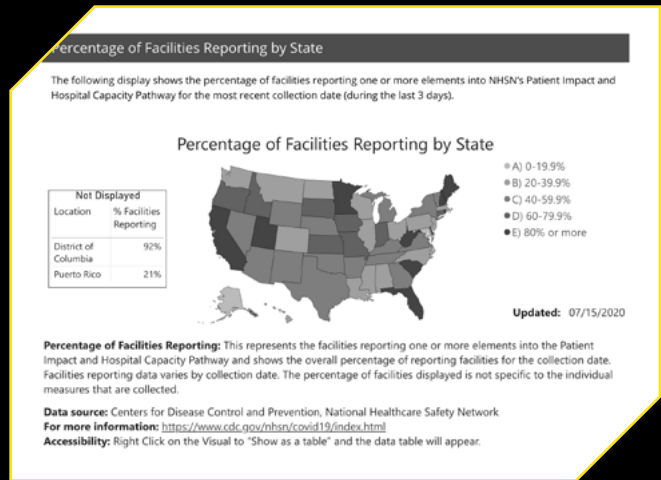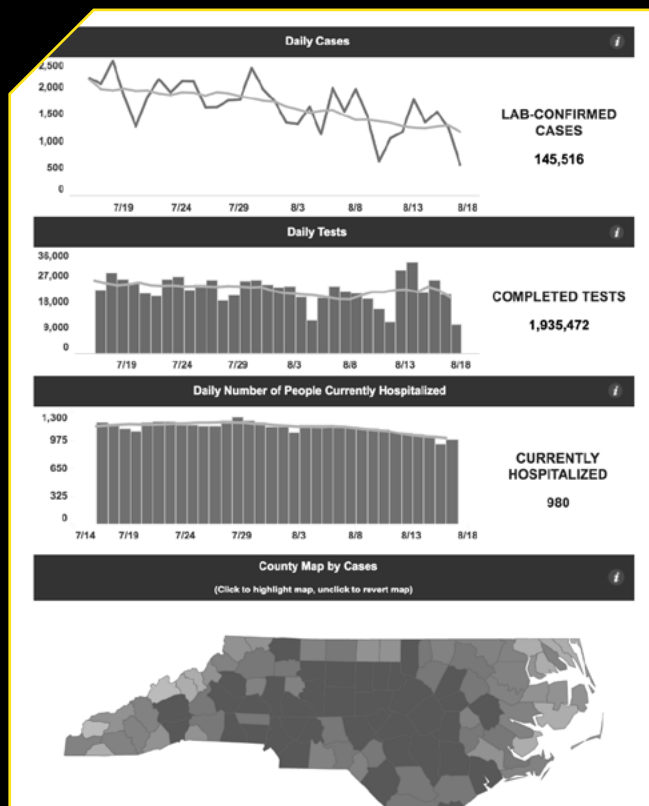
WHAT MAKES A GOOD DASHBOARD?

Simplicity.

"For a lot of the dashboards my team creates, we want to know just how something's changing. We want that quick snapshot of data that when you look at that dashboard, you can tell what's going on. **So all dashboards, for them to be really valuable, they need to be consumable and very, very simple.**"

— John Correllus, Chief Data Officer, North Carolina
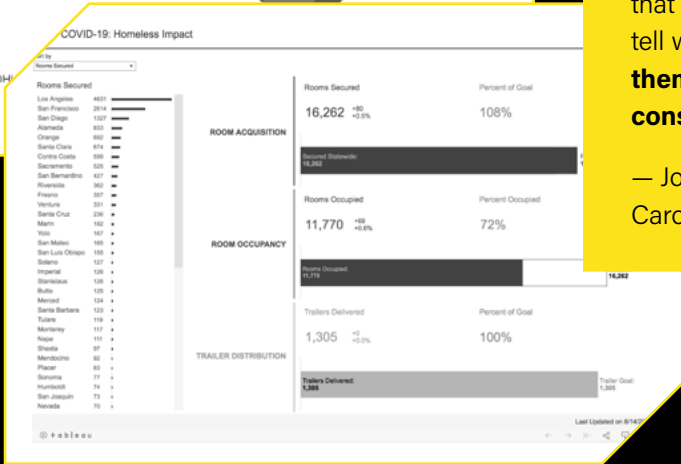
Clockwise from top left: COVID-19 North Carolina Dashboard, CDC's COVID-19 Module Data Dashboard, California's COVID-19: Homeless Impact Dashboard
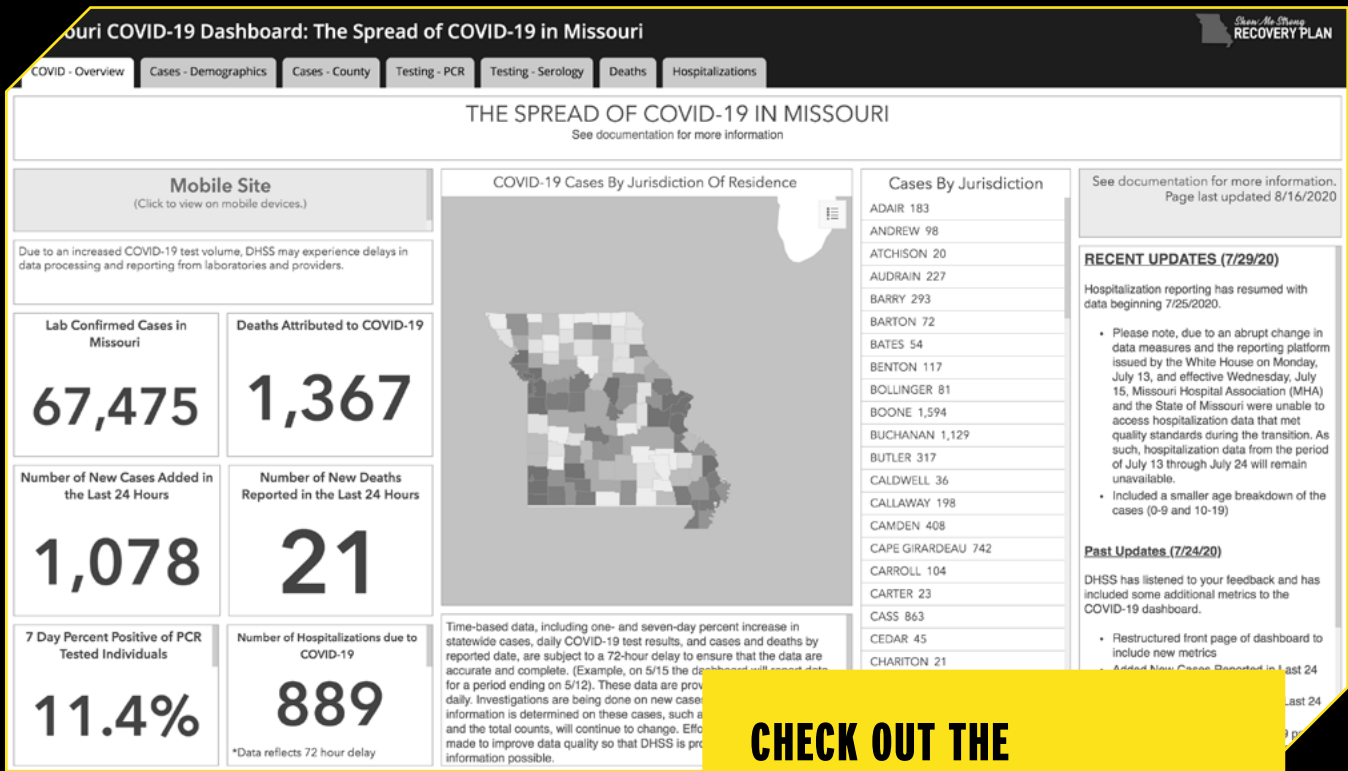
Above: Missouri COVID-19 Dashboard

Right: Oregon Health Authority COVID-19 Dashboard
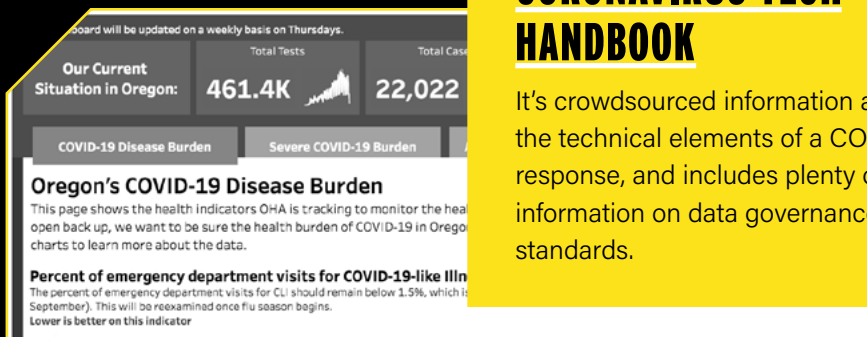
CHECK OUT THE CORONAVIRUS TECH HANDBOOK

It's crowdsourced information about the technical elements of a COVID-19 response, and includes plenty of information on data governance and standards.

FACT OR FICTION?

The seasonal flu kills as many people in a year as COVID-19 has.

Fiction: The flu was associated with 34,200 U.S. deaths in 2018-2019, according to CDC estimates. The coronavirus has killed well over 180,000 people in the U.S. as of August 2020. The seasonal flu fluctuates in seriousness.
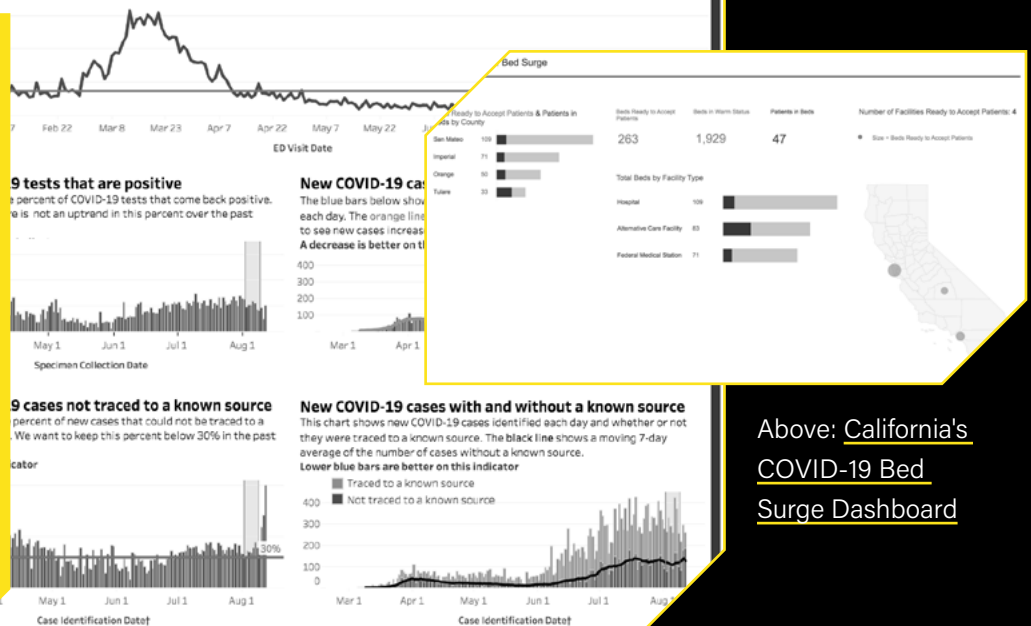
Above: California's COVID-19 Bed Surge Dashboard

For an open relationship with your data.

DataDater

Business User

⊙ 3 miles away   ● Online now

💬 Send a message

Eager overachiever seeks data with benefits.

Insights are great, but now I'm on the prowl for some real action. Looking to connect with all of my data, do a little exploring and experimenting, and finally get down to business. 😉

Looking for:

Smarter Decisions   Optimized Processes

Digital Transformation   Accelerated Value

Qlik Q ®
LEAD WITH DATA ®

Multi-cloud Data Integration & Analytics | qlik.com/us/publicsector

ANALYTICS LESSONS FROM EVERY LEVEL OF GOVERNMENT

An interview with Heather Gittings, Global Public Sector & Health Care Director, Qlik

COVID-19 is different from prior disasters in that the pandemic struck every function of government, throwing traditional response plans out the window and demanding creative, rapid-fire solutions to unthinkable challenges.

In practice, the pandemic served as a universal trial of agencies' preparedness and modernization. The verdict? While no agency was totally ready, the ones quickest to their feet had widespread data literacy and readymade use cases.

"If there's a silver lining here, it's that data really has taken center stage," said Heather Gittings, Global Public Sector Director at Qlik, a data software provider for government.

Gittings shared with GovLoop the stories and lessons of several agencies that responded effectively to the pandemic.

DATA SHOULD ALWAYS DRIVE DECISIONS.

Unsurprisingly, success stories early on came from agencies that were advanced in their analytics journey.

Early in the pandemic, the Defense Logistics Agency (DLA) partnered with the Office of the Undersecretary of Defense-Comptroller (OUSD-C) to support governmentwide response efforts with the Defense Department (DoD) supply chain. Using the advanced analytics ADVANA platform, DoD identified surpluses of personal protective equipment – including 20 million N95 respirators – to send to front-line agencies.

"They were able to act fast because both DLA and OUSD-C were already quite advanced in using data to deliver on their mission. They had access to the data, and the analytic capability to gain insight from that data," Gittings said.

EMPLOYEES SHOULD UNDERSTAND DATA.

Data-involved decisions often reconcile many different factors. Schools faced this challenge when transitioning to distance learning.

Loudoun County, Virginia, anticipated some of the difficulties that would come with distance learning. One such issue was making sure that disadvantaged students had the internet access they needed to attend online class.

The county found it had enough money to purchase and distribute 1,500 Wi-Fi hotspots, Gittings said. Instead of using one factor to determine who should get the resources, however, the county took a triage approach, using a matrix of grade, income and number of students in the household to maximize utility.

"There is a component of how to understand the data, how to read the data, how to argue with the data," Gittings said.

AGENCIES SHOULD LEARN FROM OTHERS.

The examples of heroic government responses are endless.

The National Health Service in the United Kingdom used population health data to track who was most at risk and where bed capacity was perilously low. The Waitematā District Health Board in New Zealand used contact mapping to limit community spread.

The point is, the model for success is out there, Gittings said. Now, agencies need to create the structures internally to apply the lessons others learned.

Industry partners like Qlik, which in addition to analytics offers data integration solutions, can help agencies get their data in a place to replicate these use cases. Now is the ideal time to make the transformation, Gittings said.

"If you're working to solve a problem, the chances are you're not alone. The more we can share across governments, the more we can foster success," Gittings said.

A PICTURE IN TIME:
HOW NIH HAS RESPONDED TO THE CHALLENGE OF COVID-19 DATA

Every week, Dr. Susan Gregurick changes her videoconferencing background to a favorite photo from vacations past. The transposed digital image is a moment of pre-pandemic serenity for the Associate Director for Data Science at the National Institutes of Health (NIH). Since the start of the pandemic, Gregurick has been working seven days a week to direct and coordinate data efforts related to the many COVID-19 studies, clinical trials and tests under NIH's umbrella.

On the day of her interview with GovLoop, Gregurick's backdrop is a restaurant in Sanibel Island, Florida, whose walls are inch-by-inch spattered with an eclectic assemblage of framed photos and artwork. From the trip, Gregurick remembers that the auspicious décor captivated her children – young at the time.

"I have two kids at home [now], and they're both young adults," Gregurick said. "But you know, they're only going to tolerate this life for so long before they themselves get a little antsy."

In addition to the everyday COVID-19 challenges of balancing family, pets and technological needs in her remote setup, Gregurick is also tasked with coordinating the data activities of eight coronavirus-focused NIH teams, each with 50 or more dedicated NIH staffers. Those teams have special COVID-19 focuses, including developing rapid diagnostics, creating data infrastructure for near-to-real time clinical data and understanding multisystem inflammatory syndrome in children.

One such team is the Rapid Acceleration of Diagnostics (RADx) program, which aims to speedily scale up the number of available tests and make them faster, easier to use and more accurate. The RADx teams award grants to academic labs and companies, a familiar practice for NIH research.

The NIH approach to managing data during COVID-19 is based on a hub-and-spoke model, Gregurick said. Data is passed by awarded projects – the spokes – to higher-level data coordination centers – the hubs – until it reaches a data aggregator at the top. In the case of RADx data, the final hub is the COVID-19 RADx Data Hub, which is a data aggregator, and on the way, data undergoes quality assurance and control. This approach aims to make COVID-19 data available to the research, scientific and medical communities.

"Coordination centers basically serve as a hub for data and information, and the grantees are the spokes and they feed that data in," Gregurick said. "Those coordination centers then become a spoke for a bigger hub. So, it's sort of building up concentric circles of information."
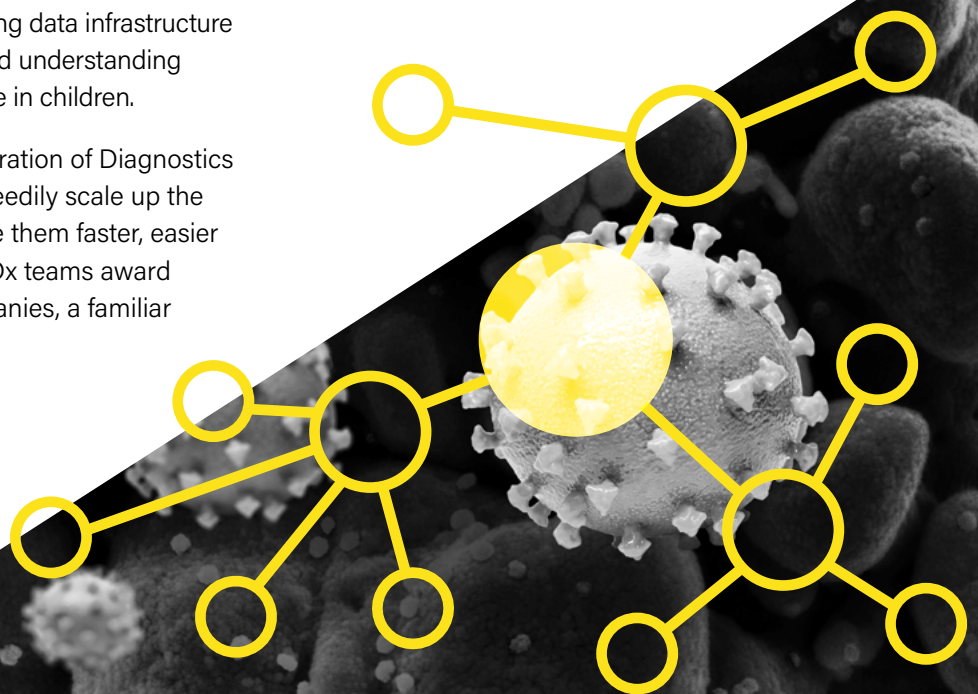
The advantage of NIH's hub-and-spoke model is that data can be looked at from a highly focused or bird's-eye view. Therefore, NIH researchers can collate COVID-19 data for information on very specific demographics, such as lung capacity in teenagers, or examine the broader health information of participants who've been tested.

The reason NIH can aggregate and analyze so much data from different sources is the standardization and frameworks put into place before data is ever input. Grantees must meet data use, standards and sharing requirements that the teams define. Privacy standards are being set up, Gregurick said.

Gregurick has helped lead a trans-NIH working group to establish common data elements. With these foundations, NIH can use algorithms to identify the profiles of individuals who might track across different systems. Doing so is important to understanding the many different health factors that COVID-19 impacts.

"If you're Jane Doe, and you're in an All of Us study, and you're also in a clinical trial, we can identify you as the same person, and therefore we can make sure that the data is linked to that one person," Gregurick said.

FACT OR FICTION?

U.S. health care providers all use the same data source to log and track patient information.

Fiction: While health information exchanges (HIEs) exist across the U.S., they are federated, and some health care providers do not use HIEs.

COVID-19 came to the U.S. with plenty of unknowns. In March 2020, many within the medical and research communities mistakenly believed that COVID-19 was only dangerous to elderly populations and just impacted the respiratory system. At the federal level, major testing and research initiatives often trailed the virus's arrival by months, such as when the RADx program began April 29, 2020, following Congressional appropriations.

In a climate of continuing uncertainty, one challenge NIH faces is that data repositories are focused on specific diseases or data types, complicating data-sharing for COVID-19, which has a wide range of health effects across populations.

That problem has been compounded because data-sharing agreements were not in place before COVID-19, a challenge many agencies encountered during the pandemic.

"The real problem is that when you create a data platform that's very specific for a particular mission, you don't necessarily think about the transaction of all the different data, like aging and child health, and heart, lung and blood missions that could be a part of COVID," Gregurick said. **"COVID makes you realize that it's a disease that has many different important contributors."**

Since initial delays, NIH research has included focuses on COVID-19's impacts on the heart and blood, as well as studies of other demographics, including mothers and children.

Gregurick pointed out that all of these efforts are driven by a common mission: to save lives. While with a vaccine and the right response, the pandemic itself will fade, its long-term health impacts will live with those who contracted and survived the virus.

Interoperable, nuanced data will be vital to treating their conditions.

"We'll be out of this soon, hopefully. We'll be out of this particular situation. But the people who've gotten sick and who've recovered may have lifelong or some periods of time of health issues," Gregurick said. "We do need this data to understand their health."

alphasix corporation ®

Hewlett Packard Enterprise

AlphaSix Corporation is a data management company. Our mission is to provide targeted IT solutions to the Federal Government that enables management of data and IT infrastructure while improving operations, understanding threats, and reducing costs.

Learn more at alphasixcorp.com.

Big Data & Analytics

Cyber Security

Cloud & Enterprise Architecture

Virtualization

Managed Print Solutions

Read our new report: *Big Data: Accelerating Time to Mission-Critical Insights*

Watch our free online course: *Extracting Mission-Critical Knowledge From Big Data*

WEAVING DISCONNECTED SOURCES INTO A DATA FABRIC

An interview with Stephen Moore, Chief Technology Officer, AlphaSix

Government agencies have turned their attention to limiting the trafficking of opioids and prescription drugs — an issue known as drug diversion. For such a pressing national problem, it seems like cities, states and federal agencies would be working in chorus.

But as those familiar with public sector data will recognize, helpful information is often inaccessible, practically speaking, because it's dispersed across data sources.

"Agencies are getting only a small, narrow view of the information they need to tackle the problem of drug diversion," said Stephen Moore, Chief Technology Officer at AlphaSix, a service provider that focuses on the convergence of big data and cybersecurity.

One emerging solution is to create a cohesive data framework for integrating and analyzing data, so GovLoop asked Moore for a few tips on how to do so. Just discovering use cases, such as applying big data to the opioid crisis, is a good start, he said.

BUILD A DATA ANALYTICS FRAMEWORK

What single sort of infrastructure or policy can support a rapidly changing field like data? The short answer: There is none.

"There isn't a big data silver bullet," Moore said. "You have to tie together the infrastructure, systems and security with a flexible analytical framework."

What Moore speaks about – the infrastructure, systems and security on top of a flexible analytical framework – is known as a data fabric. A data fabric is a series of services and systems that connects data and coordinates its management across an enterprise. AlphaSix has a data analytics framework with a combination of underlying technologies that lets agencies intake and analyze broad amounts of data.

If that sounds like a big commitment, it is. Moore said a data fabric requires support from the top brass, who suffuse it throughout an enterprise.

COLLECT AND LOAD DATA SUSTAINABLY

Data collection is not a one-time occurrence. Agencies collect data on a regular basis, and how they do is always changing – as are the data types they collect.

Some data is structured, some is unstructured, and some is a mix of the two. Without a data analytics framework, all of this information stays in its respective pods. But after stitching a data fabric together, these data sources begin to interact in a clean, structured pool that experts can draw from.

RESOLVE ANALYSIS INTO INSIGHTS

Once big data is clean and compiled in a central framework, agencies get to the fun part. They can now layer on machine learning, visualization tools or queries and filters.

At this stage, agencies really start to solve the use cases they identified out front. The analysts spot trends and pinpoint anomalies.

In the example of drug diversion above, they could pick out doctors who prescribe inordinate amounts of opioids and identify neighborhoods with higher rates of use.

"Even more important is for agencies to see the big picture of what big data can do," Moore said.

DATA TRENDS DURING COVID-19

Derived from experts directly involved with the COVID-19 response, these trends stand out as common experiences for government agencies. Funding, prior data investments and clearly defined roles and policies were integral to successful responses.

HEALTH INFORMATION EXCHANGES EXPAND IN RESPONSE TO COVID-19

Data-sharing saves lives. The COVID-19 death toll is a number so numbing that it's easy to gloss over the fact that behind every life lost, health care providers are frantically trying to treat patients with unknown conditions and diverse maladies.

Seconds, minutes and hours matter on the front lines. For that reason, health information exchanges (HIEs) save lives by expediting the time it takes to diagnose and treat patients, and North Carolina's NC HealthConnex program has been perfectly suited to the COVID-19 medical response.

"It's really about building ecosystems across the country," John Correllus, Chief Data Officer for North Carolina, said.

NC HealthConnex and HIEs generally work as neutral data stewards, providing a centralized store of medical data that hospitals and health care facilities securely tap into for patient records. The technology saves valuable treatment time by providing health professionals with patient information that they otherwise would have to test for or diagnose on site.

If a patient has a pre-existing condition, for example, COVID-19 can be especially dangerous. Health care providers can better respond to COVID-19 when they know patient information, but information silos within and across health care units can bottle up pertinent patient data.

HIEs compile data from disparate sources into one hub of information that health care providers can access on the spot, as long as they're part of the network. North Carolina has expanded data feeds during the pandemic, and about 120 hospitals are currently part of NC HealthConnex.

"If COVID would have happened even two years from now, we'd have been even more connected, so that's one of our top priorities," Correllus said.

COVID-19, Correllus believes, has highlighted systemic issues with clinical data-sharing and shown that the country needs a national data-sharing ecosystem for patients.

In the U.S., patients' information is not kept in one place. While hospitals all maintain patient records, the information that one hospital has, another care facility might not. Therefore, patient information needs to be

transferred – that is, if it can be – whenever someone switches providers or needs more immediate and urgent treatment. But time is not always on the side of those whose health care records are sealed off in another facility.

An American Hospital Association report found that 70% to 80% of hospitals have systems in place to share information with other hospitals and ambulatory care facilities. Technical and operational challenges persist, however, and large hospitals are more likely to share information.

For a variety of reasons, North Carolina has had fewer COVID-19 deaths than less populous states, as of July 2020. NC HealthConnex is a contributing factor.

But Correllus and his team are not resting on their laurels. A statewide mandate from the governor's office requires them to integrate more health care facilities as the year passes.

"We wished it was even further along than where it was today. We've made great strides and garnered great support … but again, we wish we had more of the state connected," Correllus said.

The data from HIEs isn't only being used on the front lines. North Carolina is also coordinating with public health agencies to look at aggregate trends reflected by HIE-stored data. During the pandemic, the state would like to dredge up these insights to create better policy and inform decisions.

"There's going to be a lot of policy questions that need some good data analysis and research even after we're through with this pandemic," Correllus said.

How do health privacy laws and regulations come into play? **This blog** explores how North Carolina counties can navigate confidentiality and privacy laws for COVID-19 data.

"Those who cannot remember the past are condemned to repeat it," wrote Spanish-American author George Santayana. Mistakes made during initial COVID-19 responses can be corrected with the right information and research. Economic and health researchers depend on high-quality, accessible data to understand the pandemic's impacts and inform policies going forward.

VERITAS™

Improve Resiliency in the Era of More

Address IT Complexity with the most versatile data protection platform on the market

Learn more at **www.veritas.com/resiliency**

IMPROVE RESILIENCE IN THE ERA OF MORE

An interview with Barry Levine, Director of Federal Health Care and Civilian Strategic Programs, Veritas

Recent events have been a whirlwind, but agencies aren't out of the storm just yet. Even after the COVID-19 season passes, communities will still find themselves in tumultuous times.

That's where resilience kicks in.

"It all comes down to an agency's most critical asset: its data," said Barry Levine, Director of Federal Health Care and Civilian Strategic Programs at Veritas, which specializes in enterprise data solutions.

Merriam-Webster defines resilience as the "ability to recover from or adjust easily to misfortune or change." And believe it or not, data largely determines organizational resilience. If agencies have the information at hand to make decisions, they can successfully anticipate and respond to challenges.

Agencies need a plan for resilience – specifically, a data strategy. After the Federal Data Strategy launched in 2019, many federal, state and local agencies took it as inspiration to develop their own. In the ongoing COVID-19 climate, that realization has become even more critical. When drafting a data strategy focused on resilience, it's key to keep in mind:

1. VISIBILITY

Visibility makes security, analysis and compliance easier. The goal, Levine said, is a single-pane-of-glass view so that agencies can explore all of their data repositories at once.

For some agencies, a majority of their data is dark – meaning they don't know what it is.

"Step One, get visibility into your data," Levine said. "Understand what data's out there and where it resides. Find out if it's useful."

By gaining visibility into data, storage and backup infrastructure, agencies take control over data-associated risks and IT infrastructure. No longer can they be unaware.

2. AVAILABILITY

If data isn't available, it's useless. Agencies can't protect it, use it or analyze it.

Even more than unlocking insights, agencies with highly visible and available data can find information faster and are more resilient. But just as important as using valuable data is getting rid of unimportant data. That data does nothing but drive up costs and add risk, especially in the cloud.

"Having a long-term retention strategy before it moves to the cloud could save an agency millions upon millions of dollars," Levine said.

Veritas has a classification engine that alerts agencies to security or compliance risks for their data. It can also assign different policies for various datasets.

3. PROTECTION

Cybercriminals attack governments more during crises. Nowadays, preparedness requires going beyond simple backup and recovery to enterprise data protection, Levine said.

Traditional backup and recovery solutions can fail in several ways. First, they might not capture all recent data within a few days' period, which would be disastrous if a major application is impacted. Second, data may take weeks to be recovered fully, if it even can be. For many services, such as housing applications and tax payments, "weeks" is not a reasonable timeframe.

Enterprise data protection speeds up recovery and safeguards backups across systems, which is especially important with many agencies placing data in the cloud. Agencies can even test their systems without going offline.

"Start working on a unified enterprise data protection strategy," Levine said.

GOVERNMENTS JOIN FORCES IN THEIR RESPONSE

Nobody was ready for the coronavirus. So, as COVID-19 came crashing down on the U.S. like a wave, first striking the West Coast before spilling into the rest of the country, state and local governments relied on one another to fine-tune their responses.

The State Chief Data Officers Network, housed at Georgetown's Beeck Center, is one such partnership through which states exchanged best practices, resources and information. The open market of ideas helped states plan out their responses and move in stride with or in advance of the crisis, Tyler Kleykamp, Director of the State CDO Network, said.

In these partnerships, states that first noticed COVID-19 spikes – such as those on the West Coast, where the virus was initially discovered on U.S. soil – informed governments in other parts of the country of the steps they took, such as constructing dashboards and establishing data-sharing agreements. Because of the warning, other states had a few extra days to prepare for the surge of cases.

"A couple of days means saving lives," Kleykamp said.

One common challenge early on was tracking hospital capacity once COVID-19 began to overwhelm facilities. Whereas many states dealt with hospital shortages of beds, they often felt the swell at different times.

Kleykamp noted that many reporting systems were set up to track the number of beds but not the overall capacity. In other words, states only knew the numerator, and not the denominator. These challenges, when caught and addressed on the front end of responses, helped public health authorities brace for influxes and arrange for aid, such as bringing in the National Guard to construct temporary field hospitals.

The National Guard also used data from state and local governments to prepare its response. The Joint Artificial Intelligence Center (JAIC), the military artificial intelligence branch, started Project Salus to prevent shortages in regional supply chains.

With data-fueled artificial intelligence algorithms, the JAIC predicted which places would soon be overwhelmed. Then, the Defense Department (DoD) sent necessary supplies and personnel.

"In the height of the pandemic, we were able to deploy a lot of our medical personnel, as well as other job functions that the military could provide, to help New York City respond to the pandemic," said Sunmin Kim, Chief of Policy at the JAIC.

Elsewhere, volunteers stepped up to help local and state governments meet COVID-19 demands. The **U.S. Digital Response** (USDR) is a group of 5,500 unpaid volunteers that answers to governments seeking technological aid during the pandemic. Founded by former U.S. Chief Technology Officers, USDR has taken on more than 100 projects across about half of U.S. states, as of July 2020.

"One thing that I think is really awesome and exciting about our work is we do everything open source, do everything free and, from Day One, when we build something for one city or state or county, we try to make sure it's customizable for anyone else," said Raylene Yung, CEO of the U.S. Digital Response.

Initially, Kleykamp and Yung noted, state and local governments' most dire data need was coordinating health information. Tracking available beds, creating dashboards, and identifying spikes and trends were priorities.

As the pandemic raged, state and local governments largely constructed well-working systems for public health data. But then finances followed, with surging unemployment and reactive economic assistance programs.

After the CARES Act made grants available and distributed funding to state and local agencies, governments had to account for new funding streams. In response, USDR built a tracking and alert system so that Rhode Island and Ohio could capture grant opportunities and funding.

The particular solution was a web-scraping tool that users could configure to find certain search terms. The automated solution then populated a spreadsheet showing all the available grants that matched the desired criteria, Yung said.

Broadly, the value of cross-governmental technology partnerships is that members can diagnose larger trends and needs – an essential part of pandemic responses.

For that reason, USDR built open source software that only had to be tweaked minimally across agencies to work, and the State CDO Network put out best practices and resources based on the experiences of its members to assist other agencies along the way.

"There's going to be a pretty hard look at, across state and federal government, how data was used and leveraged," Kleykamp said.

Interested in improving your state or local government's response to the pandemic? Click here for seven best practices compiled by the State CDO Network.

LOCATION DATA IS GUIDELIGHT IN RESPONSE

Hard to miss in the data rush of the pandemic has been the explosion of geospatial information.

In the past, state and local agencies especially depended on geospatial data to coordinate community programs, emergency responses and location-specific services. **But to combat the spread of an infectious disease, which requires contact tracing, geospatial information has shown what feet on the ground or eyes in the sky can't.** Contained in GIS, location data is powering responses and targeting resources to the most impacted areas.

"The geospatial community understands the value of authoritative data and high-quality data and data with standards," said Kathryn Helms, Oregon's Chief Data Officer.

Helms said when Oregon created her position in January 2019, the state geospatial team had already established standards and identified data elements. These practices had earmarked wildfire mitigation as a data-led effort before COVID-19.

Geospatial data is also crucial to creating dashboards that break down information by location. One of Helms' staffers is using data about COVID-19's spread at community nursing homes to create dashboards for the Oregon Health Authority.

"We've seen how critical location is to this and understanding deeper issues," Tyler Kleykamp, Director of the State CDO Network, said.

With GIS, agencies can analyze deeper issues. Whereas raw data shows that Black and Latino people have had significantly higher rates of COVID-19 than other populations, it doesn't reflect the reason. GIS can give the case-specific why, offering information down to the block to pinpoint hotspots.

Geospatial data could, for example, uncover an overcrowded nursing home that spread the outbreak in one of those neighborhoods. Then, agencies could go about remedying the issue by providing more sanitation resources or assistance.

"People are starting to make assumptions about comorbidity based on race. But really when you unpack that a little bit more, and look at more granular geographic areas, it might be there's a lot of public housing in that area," Kleykamp said. "Or there might be more essential workers in a particular area that are disproportionately around communities of color."

Kleykamp said he hopes states put more emphasis on mapping going forward to defeat intractable problems that outlive the pandemic.

Geospatial data has been crucial to targeted state and local COVID-19 responses. It can also be a boon in planning for and reacting to natural disasters and other unexpected events, as well as discovering disparities in funding and resources.

RESPONSES WERE ABOUT THE FUNDAMENTALS

Successful data responses to the pandemic weren't because of fancy technology like artificial intelligence. Instead, the main differentiator separating successful governments and unsuccessful ones comes from longstanding data fundamentals, experts agreed.

Data inventories, sharing agreements, standards and governance models underpinned governments that were proactive and adaptable in their responses.

"A lot of this is about investing in the foundation," John Correllus, North Carolina's Chief Data Officer, said.

As agencies grapple with the volume and value of data, they need to get their houses in order, experts said. That sort of coordination can come internally but usually requires some sort of formalization.

Kathryn Helms, Oregon's CDO, is of the belief that data stewards already exist within organizations that lack defined data departments – a philosophy called **"noninvasive data governance"** that she attributes to Bob Seiner, an author and educator in the data governance field. Those are the go-to data people, she said, and many already have their own standards and agreements in place.

Still, other efforts require more coordination and enterprise data governance. That's why Tyler Kleykamp – when he was the Connecticut CDO, before his current position as State CDO Network Director – and Helms separately drafted state data strategies for Connecticut and Oregon respectively, both of which were modeled off the Federal Data Strategy.

One priority that the Federal Data Strategy and statewide equivalents emphasize is data inventories. In fact, cataloging data is one of the first action items in the Federal Data Strategy.

Government data experts have advocated for open data, which improves transparency and the ability to share data and code.

"What emergencies show, whether it's hurricanes or pandemics, is you don't always know what issue is going to come up and how data can support that," Kleykamp said. "Having an inventory of data assets so that when an issue comes up, you know what data you have ... was one of the more critical things we heard."

Strategies that prioritize standards and catalogs are often drafted out of central data offices. However, that's something many states still lack, and comparatively few local governments have the resources for.

Jill Suurmeyer, Research Analyst for the Association of Minnesota Counties, said most counties in the state don't have employees focused on data. For all the smaller counties, it's just not something they can afford.

Suurmeyer herself compiles diverse data – such as demographics, expenditures and opioid usage – at the county level and gathers it in a "data pile." She took the idea from the California State Association of Counties, and while the data is not coded or formatted specially, the data pile is a collection of raw, usable information for counties that might not have the time to retrieve it.

For many of these smaller counties, having a data office is not a real option. Still, county administrators have taken an interest in online learning programs so that employees are familiar with the basics of data and coding. That trend, Suurmeyer said, was catching on before COVID-19 and has only accelerated since.

The data playing field is uneven. Doing what you can might start with populating a spreadsheet and looking for any figure that sticks out.

"It was less about creating a data office and more about training your current employees to be better analysts," Suurmeyer said.

Kleykamp still considers a large share of his current position to be advocating for the establishment of chief data officer roles at the state level. Several states established CDO positions during the pandemic, but others still lack the position entirely.

CDOs are central to establishing standards, data-sharing agreements and partnerships.

"A lot of discussion is around standards, and [the] kind of agreement," Raylene Yung, CEO of the U.S. Digital Response, said. "If you think of using a common standard for tracking positive cases and hospital beds and transmissions and contact tracing data, if that all ends up being standardized, now neighboring counties or neighboring states can really quickly get a better sense of the overall picture."

Yung and Kleykamp recommend that states also build on open data and principles, which are datasets and code available to the public.

Still, when evaluating responses, governments must be graded on a case-by-case basis. Small local governments across America are still working off fax machines and Microsoft Excel software that is two versions too old for modern needs.

"A number of counties didn't have the most current version of Excel," Suurmeyer said, illustrating the technological constraints of some county governments in Minnesota.

Akamai

Experience the Edge

IT Modernization Encourages Moving Your Critical Applications to the Cloud

The surge in your agency's remote workforce demands secure access and availability of those applications at all times.

A Zero Trust architecture ensures that when doing this, you trust but verify.

Learn more about how Akamai can partner with you on your journey to Zero Trust.

IS A PERIMETER SECURITY MODEL PROTECTING YOUR DATA IN 2020?

An interview with Patrick Sullivan, Chief Technology Officer of Security Strategy, Akamai

The year 2020 has highlighted the many problems of a perimeter security model. As employees and contractors work remotely, applications leave agency data centers, leading adversaries to exploit the soft underbelly found inside the perimeter.

Security inspection is moving from the data center to the edge, closer to employees and applications. Agencies now have the opportunity to optimize.

"Any assumption that a user should be trusted because they're on an agency network should be challenged," said Patrick Sullivan, Chief Technology Officer of Security Strategy for Akamai, a global cloud security company.

Agencies have traditionally operated off the assumption that if the perimeter is secure, their data is too. But in a distributed environment, that isn't necessarily the case.

In an interview with GovLoop, Sullivan shared several tips for how agencies can transition to a secure remote environment.

SEE FLAWS IN THE STATUS QUO

The perimeter security model is failing to protect agencies' most critical data, Sullivan said. Once inside these architectures, attackers can consistently exploit the model with easy lateral movements throughout the network, often guarded by virtual private network (VPN) connections.

But were VPNs really meant for the challenges of 2020, when multiple devices use the same connection and multiple family members use the same device? No, Sullivan says.

"Lacking office controls in unpredictable environments, that further underscores the challenge in granting trust at the network layer via something like a VPN," Sullivan said.

Furthermore, once logged onto a VPN, employees often receive an access grant to the network that's valid for hours at a time. Without security nearby or office firewalls, that autonomy can be a vulnerability.

UNDERSTAND ZERO TRUST

Enter zero trust. Zero trust is a security architecture that increases scrutiny of access, moving from a grant to access the network to interrogation for each request. Interrogation can include checks for business needs, device posture and anomalous requests down to the application or data layer.

It works much like a hotel key card. The card authorizes access to the building, community rooms – like laundry and workout rooms – and the guest's room. It doesn't, however, open others' rooms or employee-only facilities.

Not only does zero trust limit what users can access, but it also verifies users' identity and permission levels every time they request access – in contrast to excessive permissions often given in VPN sessions.

Not coincidentally, zero trust is designed for remote work, mobile and cloud environments.

GO STEP BY STEP

Make no mistake: Zero trust is a transformation, not a tweak.

And yet, agencies can approach it incrementally. Guidance from the National Institute of Standards and Technology outlines steps for adopting a zero-trust architecture. Vendors such as Akamai can help design a plan for how to move to zero trust without disrupting business.

Agencies can even prioritize certain communities or applications for zero trust. The process doesn't need to be all or nothing.

"The big takeaway is, this isn't something that you do all at once," Sullivan said. "It's a continual set of steps and you learn from each one as you go."

PRIVACY AND TRUSTWORTHINESS STILL AT THE FOREFRONT

The South Korean government uses personal data to trace coronavirus infections by tracking credit cards, cellphones and other private devices. And while the approach has been credited with slowing the spread of the coronavirus, it has also raised serious concerns about data privacy and ethics.

The United States doesn't have a program resembling South Korea's contact tracing, but still, data privacy concerns loom. The Defense Department's Joint Artificial Intelligence Center, for example, has acquired personal information from businesses and civilian agencies for its coronavirus response program, Project Salus.

"For us to, for instance, predict how COVID-19 is going to be affecting a certain military installation, you might need data from things that would be nontraditional DoD data," said Sunmin Kim, Chief of Policy at the JAIC.

Kim said that DoD takes precautions so that citizen trust and privacy are not violated. Personal information is anonymized and geospatial data is given within a range, instead of at a granular detail, to defend against abuse. The JAIC even sacrifices modeling accuracy to protect privacy, Kim said.

"Let's say we're just talking about testing results," Kim said. "Maybe we only want to have it at a certain ZIP code or a Census tract level, versus something that could be distilled all the way down to a singular block, because that means that there's a greater risk of identification."

DoD has broad interpretation authority over what data to collect and how to use it. The Privacy Act of 1974 permits federal agencies to incorporate data critical for the mission, Kim said, meaning DoD is ultimately stamping its own clearance for what data to request. Therefore, her team at the JAIC has a lot of flexibility, and uncertainty, about what data elements can be used and to what degree they should be insulated.

Clearly define what data is being used and what its potential limitations are. Transparent processes will lead to greater citizen trust.

DoD Secretary Mark Esper published five AI principles in February 2020 that have helped guide the JAIC's approach. The five principles are responsible, equitable, traceable, reliable and governable.

Kim said that the Privacy Act requirements, though outdated, are not as important to the JAIC's data ethics as agile internal protocols. After all, she said, official data policies have a relatively short shelf life, since data standards, formats, technologies and use cases are constantly evolving. To compensate for the lack of a modern law, the JAIC makes a point to have policy, legal and product teams integrated at the start of any project, something that Kim said the center did "very well" for Project Salus.

"That gets to data minimization and responsible data use," Kim said. "I think what might need to be updated over time are the internal processes that allow us to make that determination, and that's exactly what we've been focusing on at the JAIC, versus trying to interpret that law and codify internal JAIC policies."

Privacy concerns are not only present at DoD. In Minnesota, a Blue Ribbon Council appointed by Gov. Tim Walz published a report in June 2020 recommending that **"the state should appoint a chief privacy officer position to support state practices on data privacy and sharing."** The council was made up of agency chief information officers and other IT leaders.

Concerns about privacy influence trust in government as well, Kim said. She noted that while movements to make data more open arise – those often do not apply to DoD because of the sensitive nature of its data – governments are still obligated to let citizens know what their information will be used for. That's a tradeoff the JAIC weighs before ever collecting data, she said.

"When it comes to data about U.S. persons, we the government have an obligation to our constituents saying if we're collecting data on them in order to administer certain public programs. For me as a constituent, I wouldn't expect that data to be used in an unexpected way," Kim said.

THE INTERSECTION OF RACE AND DATA

None of the events of 2020 happened in a vacuum.

The challenges universally faced by Americans were compounded by existing class and racial disparities. COVID-19 data by race, acquired by The New York Times after it sued the CDC, showed that Latino and Black people have the highest positive coronavirus test rates by far.

Latino people tested at 73 cases per 10,000 people, and Black people tested at 62 cases per 10,000 people. White people, in comparison, tested at a rate of 23 cases per 10,000 people.

The disparities are not confined to state lines or regions. Experts have surmised that this discrepancy is because Black and Latino people have lower average incomes, and thus are more likely to live in densely populated regions and work in-person service jobs.

DATA AS A LENS OF RACIAL EQUITY IN COOK COUNTY

On a July 2020 webinar, leaders in Cook County, Illinois, explained the racial and socioeconomic factors they have used to evaluate the disproportionate impacts of COVID-19, and the subsequent economic fallout, on Black and Latino communities. Cook County is 26% Latino, 24% Black and 42% white.

"The data shows that the impacts of natural disasters, economic downturns and health crises do not fall equally on all of us," said Toni Preckwinkle, Cook County Board President. "Using data science and geographic analysis,

we see where impacts are not equitably distributed and provide that information to county agencies, other government entities, nonprofits and residents to prioritize our resources where help is needed most."
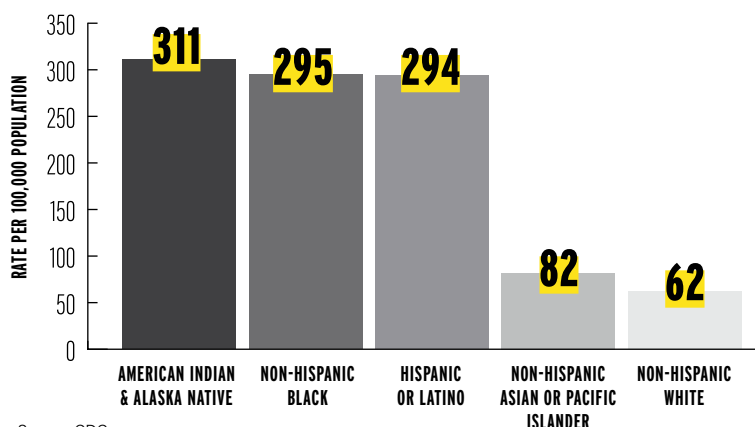
Preckwinkle, a Democrat who lives in Chicago, said that under her administration, she wants to incorporate a "racial lens" into analysis. So in Cook County, race is not out of bounds when discussing policies, programs and studies. Instead, it's a major part of public service.

"We're going to use a racial equity lens to look at how we make investments, our programs, our policy, everything," she said.

*In the U.S., Black people have died from COVID-19 at **more than twice the rate** of white people.*

See the data from the CDC here.

AGE-ADJUSTED COVID-19-ASSOCIATED HOSPITILIZATION RATES BY RACE AND ETHNICITY

RATE PER 100,000 POPULATION

Race/Ethnicity	Rate
AMERICAN INDIAN & ALASKA NATIVE	311
NON-HISPANIC BLACK	295
HISPANIC OR LATINO	294
NON-HISPANIC ASIAN OR PACIFIC ISLANDER	82
NON-HISPANIC WHITE	62

Source: CDC

That racial equity lens is shaped by data. The county uses GIS mapping technology to plot out its social vulnerability index (SVI), which tracks factors that make a community susceptible to being disproportionately impacted by crises that affect day-to-day life. By putting this information on a map, Cook County can critically evaluate what neighborhoods need more help, said Dessa Gypalo, the county's Chief Data Officer. Preckwinkle said that poor neighborhoods and Black and Latino communities have been hit the hardest.
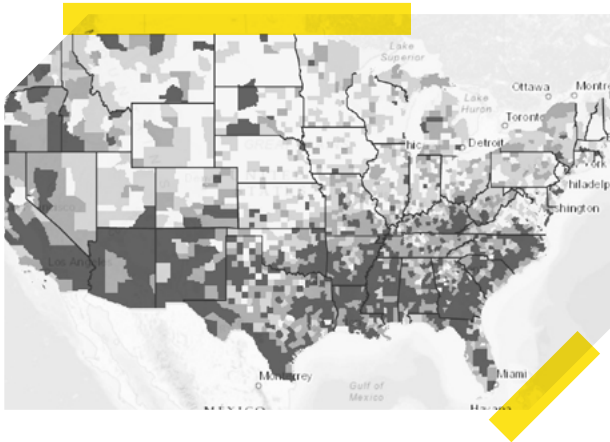
The county has used GIS for several purposes. One is for the medical examiner to pinpoint badly impacted areas. Another usage tracks investment data from different agencies to see if funds are being equitably distributed.

Cook County received about $430 million in initial CARES Act funding, and it passed along some of the money to local governments within its area. Preckwinkle said the county could have simply parsed out the funds it shared, $51 million, on a per-capita basis, but it chose not to.

In practice, that would have meant large population centers in Cook County, like Chicago, would receive the same amount per resident as smaller towns. But then, poorer communities would not have gotten the help they needed, Preckwinkle said.

"We're immediately allocating resources on the basis of need – not simply on the basis of population," Preckwinkle said. "It would have been the easiest thing to do, just to take the money up and divide it up according to the population of every city, town or village."

The county did distribute less than a third of that money, $15 million, per capita. But for the rest of its funds, it used the prism of its SVI. **The county allocated more than two-thirds of the money it shared, $36 million, on the basis of need.**

Of course, Cook County deals with its own data challenges too. During the webinar, hosted by Chi Hack Night, a participant noted that locational data about COVID-19 deaths was only available via a Freedom of Information Act request, and even then, it was "messy."

Tom Lynch, the Chief Information Officer of Cook County, also said that because of inconsistent testing methods and a disconnect between five different public health departments, location information could be inexact or unstandardized. Though location is tracked for tests, some results might report a home address while others are far more imprecise.

"There's nuance depending on case to case as to what's being reported and by which agency," Lynch said.

Still, progress is being made. Lynch said the five public health departments, which serve local communities like Evanston, Oak Park and Chicago all within Cook County, are improving coordination and data-sharing. While there was no timeline for when more fluid data-sharing could be in place, conversations are happening between local departments and the Illinois Public Health Department, he said.

Preckwinkle credited Gypalo, the county's first CDO, with expanding the vision of data usage in Cook County. Since she began, three other elected offices in the county have established CDO positions as well. Preckwinkle said that was proof of Gypalo's progress.

Gypalo, for her part, thanked her team for pushing county data forward. She said her team's top priority when looking for talent was finding people with diverse skill sets who were passionate about "actionable and shareable data." That mindset helps her office consistently find new use cases for the public good, she said.

"Really what I'm looking for, and what I think we should all be looking for, is data folks who come in and question things," Gypalo said. "They ask, 'Why?' They have very logical minds in terms of not just thinking from the start but to the end."

DATA'S USEFULNESS, AND LACK OF AVAILABILITY, TRACKING POLICE USE OF FORCE

Counterintuitively, the most comprehensive databases on police use of force often do not come from government. While local departments steward their own files, journalists and nonprofits maintain the most complete statistics on police use of force nationally.

As police use of force and racial implications drew more media attention, the FBI launched a nationwide use-of-force data collection program in January 2019. **The FBI Use-of-Force Data Collection, initially set to be released in summer 2020, includes police shootings that kill, injure or are directed at a person.**

The first of its kind from the public sector, the collection tracks race, age, sex and other identifiable characteristics, and specifies whether the subject was armed or resisted arrest. It also reports data on the police officers who were involved.

The initiative comes as racial justice groups demand more transparency, and data professionals advocate for open data. Their efforts have been underway for years.

The FBI's latest data collection is not comprehensive, however, leading to questions about its efficacy. By July 27, 2020, approximately 41% of the nation's law enforcement officers had been covered by reporting agencies. Submitting data is voluntary for agencies.

"The data collection can facilitate dialogue and educate the public concerning how law enforcement is trained on use of force," an FBI public affairs official wrote in an email to GovLoop. "Participation will promote transparency and accountability between law enforcement and the communities they serve."

Major law enforcement agencies are not reporting information, which means there will be a data void coming from some of America's largest cities. Police departments, which would have to submit the information themselves and cover associated costs, have cited resource constraints and data-sharing challenges as barriers to sending the FBI information.

The Phoenix Police Department is not participating in the FBI Use-of-Force Data Collection, though it is engaging in other transparency efforts. The city is building out its own use-of-force police database – a recommendation of the National Police Foundation – and the department will share its information with other national police partnerships including the Police Data Initiative, said Mercedes Fortune, Public Information Sergeant for the Phoenix Police Department. The Police Data Initiative is an open data project with more than 130 participating agencies, according to its website. Its information is not linked to the FBI's nascent program.

Delivering on the Promise of AI in Government

Artificial intelligence has the power to tackle the government's toughest missions. DataRobot is helping agencies unlock this power, at scale, and responsibly with the only trusted AI platform that automates and accelerates every step of your AI journey from data to mission value.

Start Your AI Journey with DataRobot Today

www.datarobot.com/solutions/public-sector/

DataRobot

SAFEGUARDING TRUST IN YOUR AI

An interview with Chandler McCann, General Manager, DataRobot's Government Solutions Team

Government officials are starting to discover the enormous potential and power of artificial intelligence (AI) in the public sector.

But building AI into an organization means redesigning the very foundation of how decisions are made. Just like when adding any new decision-maker to an organization, officials need to build a base of trust before rolling out an AI system in the real world.

Many agencies don't yet have the data science foundation to support AI programs.

"What we see is a landscape of AI opportunities," said Chandler McCann, General Manager for DataRobot's Government Solutions team. DataRobot provides a highly trusted end-to-end enterprise AI platform and supporting services. "When AI doesn't prove trustworthy, that can have drastic consequences – for government especially."

According to McCann, trusted AI starts with three critical components. Without these pillars, agencies can't trust their AI's decisions.

1. GUARDRAILS

AI can make mistakes in dozens of ways. There can be missing or mislabeled data. Models can become overconfident, or find patterns in coincidences. And without the appropriate guardrails, AI can be misled, and ultimately it can mislead humans.

Counterintuitively, you should be wary of models with perfect accuracy – as that likely means training data was too clean or included the right answers. Without knowing its limitations, your model could make an overconfident decision as a result.

DataRobot offers automated machine learning with numerous guardrails that let humans remove data that leads to false confidence. Humans can also make their own decisions when the model isn't confident.

2. GOVERNANCE

AI accomplishes nothing in a vacuum. Organizations can only reap the rewards of AI once its decisions are made and used in the real world. But these decisions have to be tracked, with someone held accountable for their success.

Just because an AI system is accurate today, that doesn't mean it will be in six months. As the world changes, so does data. And if a model is never updated, it will eventually be no better than a coin flip.

With AI governance, every decision your AI makes is tracked, so you can check it's adding value and not losing accuracy over time. Once the world changes around the model, AI governance ensures that someone is responsible for retraining the model on new data.

3. TRUST ASSESSMENT

AI depends on pattern recognition, using machine learning to find trends in data. But an AI model can learn trends from data that it shouldn't use – like race, sex or age – leading to biased or unfair decisions.

"We've seen examples where health care AI models have considered patients' race or income instead of health," McCann said.

It's possible to build an AI system in conflict with human values. But this problem is avoidable, as long as the AI includes a trust assessment to identify how its decisions affect people and ensure that its decisions align with organizational values.

AI is built to help organizations make more efficient, faster decisions. With ideal tools and proactive steps, it also leads to decisions that are fairer, more ethical and more trustworthy.

"Our goal and commitment to our community is accountability and transparency," Fortune said in an email to GovLoop.

The FBI recently made it easier for departments to report data by allowing for uploads instead of manual entry, Fortune said. But Phoenix already had established data-sharing frameworks with other associations, and the police department chose to send its data to existing connections instead, Fortune said.

The Houston Police Department is currently evaluating the cost of converting its record management system to gather FBI-requested data elements, Diana Poor, its Chief Data Officer, told GovLoop. Though it is not currently reporting use-of-force information to the FBI, the department will look for funding sources to submit data after calculating the cost, Poor said.

Houston and Phoenix are the fourth and fifth most populous U.S. cities, respectively.

Local police use-of-force information can be incomplete or inaccurate if not subject to review. A draft research paper found more than double the deaths occurred "as a result of or during an interaction with" law enforcement than what the New York Police Department (NYPD) originally reported for a five-year period. From 2010 to 2015, the department had reported 46 deaths, less than half of the 105 law enforcement-related deaths the study found. The New York Times originally unearthed the report, commissioned by the New York City Health and Mental Hygiene Department.

"More New Yorkers' deaths are due to law enforcement than are captured by our official surveillance measures," tweeted then NYC Health Commissioner Oxiris Barbot, who later controversially dismissed an NYPD request for masks and eventually resigned.

Why is there no comprehensive database for police use of force?

There's no one answer. Until recently, studies hadn't been commissioned. And even now that one has, reporting is voluntary and the decision to report data falls on local departments. The cost falls on them too, an important barrier to entry.

The report found 54% of those killed in police interactions were Black, and legal intervention deaths were "significantly higher" for Black people than white people. The authors concluded "there are significant racial disparities in legal intervention mortality, particularly among unarmed persons."

Per a later-released report, NYPD, America's largest police department, had not participated in the FBI's data collection. NYPD did not respond to a request for comment.

The goal of the FBI data collection is to offer an aggregate view of use-of-force incidents. It does not rule on whether an officer acted lawfully.

Though efforts are growing, the lack of a single, holistic open data source makes it difficult to review police use of force on a national scale.

Law enforcement agencies can still submit data to the FBI, which will report more detailed information when it has more data.

U.S. CENSUS TIMELINE

A Timeline of the U.S. Government's Oldest, Most Historic Data Collection Project
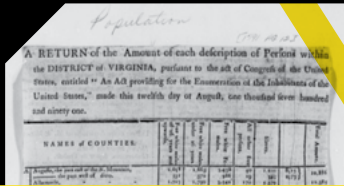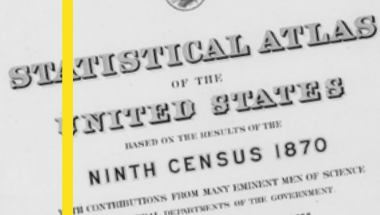
1790
The **first U.S. Census is counted**. Marshalls ask six questions as part of a population count. The U.S. population is 3,939,214.

1849
Congress establishes a Census board to oversee data collection efforts.

1872
The Census welcomes a mechanical tallying machine to expedite the tabulation of data.

1874
The Census Office publishes the "Statistical Atlas of the United States."

1902
The U.S. Census Bureau is enshrined as a **permanent government agency**.

1905
The Census collects its first inter-decennial data on manufacturing businesses.

1921
The Census moves to a biennial collection of manufacturing data to facilitate economic research and growth.

1937
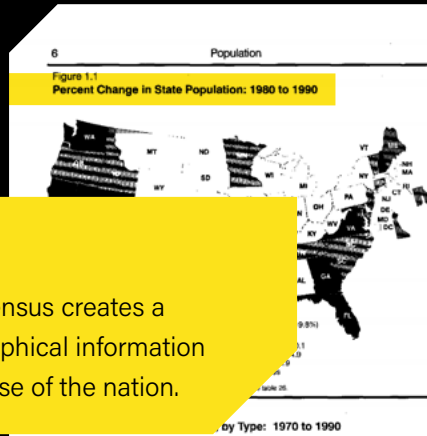The Census studies statistical sampling and conducts its first study of unemployment.

1951
The Census uses UNIVAC 1, the first nonmilitary computer.

1975
All Census data is available on magnetic computer tape.

1991
The Census creates a geographical information database of the nation.

2007
The Census releases the first "Statistical Atlas of the United States" since 1920.

2020
The bureau delivers its **first online census.**

The Technologies You Need for
Mission Success.

SECURITY

DEVOPS

BUSINESS OPERATIONS

STORAGE

DATA INTELLIGENCE

NETWORKING

COMPUTE

VIRTUALIZATION

immixGroup, an Arrow Electronics company, is a leading **value-added IT distributor** that delivers mission-driven results to the public sector through our vast partner ecosystem that includes 300+ manufacturers and 1,200+ solution providers. Together, we deliver **innovative technology solutions**, powering the future of public sector IT.

immixGroup
An Arrow Company

8444 Westpark Drive, Suite 200 • McLean, Virginia 22102
Tel 703.752.0610 • Fax 703.752.0611 • www.immixgroup.com
© 2020 immixGroup, Inc. All rights reserved. (03272020BY)

DELLTechnologies
FEDERAL AUTHORIZED DISTRIBUTOR

TEARING DOWN DATA SILOS TO CONNECT THE AGENCY

An interview with Dallas Nash, Senior Director of Sales for Global Government, Dell EMC

On the journey to modern data centers, federal agencies have undoubtedly come across the advice to "tear down data storage silos." Those are wise words, but how can agencies actually follow them?

Today's storage silos aren't just office walls. They're complicated contracts and legacy technology obstacles with proprietary lock-ins and relatedly isolated data practices.

"We have to collapse data into a central repository. To do that, you have to look at modernizing your current infrastructure," said Dallas Nash, Senior Director of Sales for Global Government at Dell EMC.

Simply, the goal is to have data live in one shared space, accessible from anywhere. That centralized repository is known as a data lake.

GovLoop recently interviewed Nash about how agencies can turn unstructured data into a collective resource. Below are the areas he outlined – the edge, core and cloud.

1. EDGE

Increasingly, data is generated outside offices and at the edge, where work is now conducted. While valuable, information at the edge is often unstructured and goes unused – envision the many video surveillance cameras' data that's never touched. Unstructured data is critically important, but lacks standards in format, type and values.

Imperative for modern data intake, Nash said, is to start with an end goal in mind. Without an unstructured data target and the ability to glean real-time business intelligence, agencies have dormant data – taking up space and driving up costs.

If the aforementioned cameras had their data pooled and analyzed to identify license plates and vehicle types, then that information carries true mission-critical value. Acted on, it could save lives.

2. CORE

For actionable information, all external sources must come together in a core, scale-out network-attached storage solution – simply, a data lake.

A data lake is a broad-scale unstructured data repository. Different data streams throughout the enterprise – like sensors, cameras and file portals – are all tributaries into one central location.

The benefit of a data lake is that data scientists can return to the same watering hole for reliable information from established sources. They don't have to waste time retrieving data from scattered, disparate sources.

"The data lake allows both customers and data creators to have mass-scale ability to pull value out of data," Nash said.

What agencies need to consider is how information traverses the network into the data lake. Vehicles like Dell Isilon and PowerScale move data from the edge to the core.

3. CLOUD

Agencies finally need to link the core – the data lake – to the cloud, where ready-to-boot artificial intelligence (AI) solutions assess and analyze unstructured data. The cloud makes insights portable and accessible.

Dell Technologies offers a low-lift AI starter pack with compute, networking and storage so that agencies can flexibly deliver insights across multiple data centers and private, public or hybrid clouds.

Eventually, agencies might permanently favor a private cloud for simplicity and consistency, Nash said. Private clouds can eliminate significant public cloud data egress costs.

"Options are wonderful, especially when looking to make that initial move," Nash said.

CDO CHATS AND TAKEAWAYS

Here are some highlights from the many chief data officers whom GovLoop spoke to as part of this report. For more information and other interviews, visit GovLoop's <u>CDO Conversations</u>.

ON DATA SKILLS ...

"I care a lot about empowerment and skill-building and who is coming on to do data analytics work. I think in the public sector in particular, we have to be very intentional about the work that we do and the impact it has, and we have to do as much as we can to reduce bias in all of the data we're doing because it is used in such a wide array of other activities."

– Dessa Gypalo, CDO, Cook County, Illinois

"I generally follow the Bob Seiner noninvasive data governance approach, which is essentially stating that you probably already have people who are stewarding data. It's like the person that you go to when you always have a question. ... So generally, those folks tend to organically emerge, especially when you have this coordination of effort."

– Kathryn Helms, CDO, Oregon

ON LEADERSHIP ...

"The primary thing is to be a data evangelist – to collaborate within and outside of the state on data issues; to integrate some of the systems within the state; to train the workforce on data technologies; and to execute on the strategic and tactical projects that we have that are involving data. There's acronyms, and this acronym I decided to call 'CITE,' so collaborate, integrate, train, execute."

– Tammy Roust, CDO, Illinois

"We oftentimes serve as translators. I use this term a lot. We serve as translators between the data itself and maybe a more technical audience with policy, operations, decision-makers, as well as the public. There's a lot to be said for how you champion not just analysis but data availability and data literacy. Using open data is one of the best ways to do that."

– Kelly Jin, Chief Analytics Officer, New York City

ON ELEVATION OF DATA ...

"Early on, it was driven by the common saying, 'Hey, the questions we're trying to answer,' and then working with CDOs to try to get answers to those. I think now we're seeing more opportunity for CDOs to advise more up the chain of command in terms of what's possible."

– Tyler Kleykamp, Director, State CDO Network

"I definitely think the importance of data has been elevated and the importance of data-sharing, the importance of cataloging your data to understand what data you have to help solve 'x' problem is very important. But I would think from a state, we have recognized the value of data in North Carolina for a very long time. And the other thing we've also recognized is an important point, but the value of data in the state increases when you integrate across the mains to develop intelligence. That's really critical that people understand that."

– John Correllus, CDO, North Carolina

ON BEST PRACTICES ...

"We're mostly knowledge workers in some way, shape or form. We either produce information, we collect information, or we create it in the course of our duties. And so, making sure that work that comes to my level is prioritized effectively so that we're saying, 'OK, we as a state are particularly committed to these core principles.' ... Otherwise what happens is you get this flavor-of-the-week reaction to data."

– Kathryn Helms, CDO, Oregon

"I used to work for [former] U.S. Chief Technology Officer Megan Smith at the White House, and she has a concept of 'scout and scale.' That means that when you are facing a challenge, go out and figure out who else is solving this problem, how they are solving the problem and figure out how to lift up those voices."

– Kelly Jin, Chief Analytics Officer, New York City

"States are going to have to make some really difficult decisions from a budget perspective and ideally would leverage data about how effective their programs are, so that they can be more surgical about how they make some of those budgetary decisions."

– Tyler Kleykamp, Director, State CDO Network

THANK YOU

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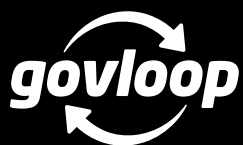
For more information about this report, please reach out to info@govloop.com.

AUTHOR

Isaac Constans, Senior Staff Writer

DESIGNER

Kaitlyn Baker, Creative Manager

![govloop](govloop logo)