



Retracing Steps

Reflecting on Management Lessons
in Public Health Data Infrastructure
During COVID-19

OCTOBER 2021



About the Partnership

The Partnership for Public Service is a nonpartisan, nonprofit organization that works to revitalize the federal government by inspiring a new generation to serve and by transforming the way government works. The Partnership teams up with federal agencies and other stakeholders to make our government more effective and efficient.

Table of Contents

Executive Summary	4
Introduction	5
Definitions	6
How Public Health Teams Adapted to Respond to the Crisis	7
Health Resources and Services Administration	7
Office of Innovation, New Jersey	10
National Center for Advancing Translational Sciences	12
California Health and Human Services Agency	15
Center for Preparedness and Response, Division of Emergency Operations: CDC	18
Johns Hopkins Coronavirus Resource Center	20
Recommendations	23
Conclusion	25
Project Team	26
Acknowledgements	26

Executive Summary

The COVID-19 pandemic is unlike other health threats seen in the past century. From January 2020 to October 2021, more than [half a million](#) Americans died from the disease. In the event of health threats like these, an interactive system of governmental and health care actors and technologies must work together to accurately detect, report, predict and facilitate responses to prevent illness and death. This system is known as the public health data infrastructure, and public health surveillance is one essential piece of this system.

Public health surveillance enables leaders to understand where a disease has appeared and how it may spread. Through strong public health data surveillance and infrastructure systems, government leaders can protect public safety by creating policies and allocating resources based on predicted disease outbreak. One [study](#) found that if social distancing policies had started one week earlier during the COVID-19 pandemic in the United States, there could have been 700,000 fewer infections and 36,000 lives could have been saved.

Early in the COVID-19 pandemic, the [federal government struggled](#) to collect accurate, comprehensive and timely disease data. Across federal and local government, data platforms and definitions were not standardized, which made data sharing across institutions and sectors difficult. Platforms lacked interoperability and fragmented systems negatively impacted continuity of care and efforts to aggregate large amounts of data. [Nonautomated and time-consuming processes](#), such as the use of paper records, faxes and phone calls to share case data, affected the accuracy and timeliness of data. Finally, a lack of data use agreements and concerns over privacy and security prevented sharing information across sectors.

Many of the challenges seen at the beginning of the pandemic still affect data collection efforts one-and-a-half-years later. States use [different definitions](#) for a positive COVID-19 test, vaccines are not being distributed [equitably](#) and policies are not always based on [public health recommendations](#). To help solve these challenges, the Partnership for Public Service conducted 15 interviews and one workshop with federal, state and non-profit leaders to identify successes stories and recommendations for congressional and federal leaders. The hope is that this information will lead to substantial and rapid progress on modernizing the nation's public health data infrastructure to improve the timeliness, quality and coordination of public health data.

This report highlights six case studies that illustrate valuable lessons learned and best practices for creating a stronger public health data infrastructure. Federal, state and academic leaders focused on workforce capabilities, technological platform decisions, data transparency and tips for creating policy.

Five key recommendations for federal leaders arose from this work:

- Invest in core public health systems.
- Ensure timely, accessible and secure data.
- Highlight public health experts as leaders.
- Standardize and define terms.
- Developing a robust, flexible and agile data workforce.
- Strengthen cross-governmental collaboration through skills trainings, report streamlining and more.

As new COVID-19 variants arise, it is more crucial than ever to invest in strong public health data surveillance and research, and to enact best practices. The [Centers for Disease Control and Prevention](#) received [\\$500 million](#) from the CARES Act to improve surveillance. Cross-sector partners and the federal government are well-positioned to create more robust partnerships, systems and policies to create a strong public health data infrastructure and save lives.

Introduction

In the event of a public health threat, an interactive system known as “public health surveillance,” comprising government, community and health actors, typically responds to drive technology-enabled detection, reporting, prediction and intervention, and prevent illness and death. This system has saved countless lives in past public health threats—from E.coli-contaminated lettuce to influenza epidemics to the opioid crisis. For the system to work effectively and improve health outcomes for everyone, it relies heavily on a robust underlying [public health data infrastructure](#) among cross-sector entities: government public health agencies at the federal, state, local, tribal and territorial levels, private health care providers, nonprofit organizations and the public. These public health data systems, however, are far from robust, and that has dire implications for this country’s most vulnerable populations.

While challenges existed long before the coronavirus pandemic, the crisis has revealed the true extent of shortcomings in the antiquated public health data infrastructure in the United States. Fragmented, [nonstandardized](#), [time-consuming and error-prone](#) processes—including the use of paper records, faxes and phone calls to share data—impacted communication and the accuracy and timeliness of COVID-19 response efforts. Data infrastructure lacked automation, [sufficient data privacy](#) and security measures, and interoperability for real-time data—features needed to be able to understand, predict and manage the virus. All of this affected the quality of the data, which often had critical gaps in demographic information on race and ethnicity, further hindering our government’s ability to develop well-informed and equitable interventions. As of September 21, 2021, more than one-and-a-half years after the pandemic started, CDC reported that race and ethnicity data was only known for about [59%](#) of people who got at least one vaccine.

Many of these technical considerations were well-documented before and during the pandemic, and both Congress and [federal agencies](#) such as [CDC](#) have begun the enormous task of modernizing public health data. However, it is not always clear how this modernization should occur. It is essential to modernize management of technologies and people. Proper governance around open data, an agile and skilled workforce, and strong partnerships between federal and state health agencies are just a few examples of the possibilities.

As congressional and federal agency leaders continue to respond to coronavirus variants and consider investment strategies for advancing public health data infrastructure, it is essential to consider on-the-ground management lessons public health teams learned during the pandemic. Through a series of 15 interviews and a workshop with federal, state and nonprofit public health leaders and practitioners, this report highlights examples of how cross-sector teams navigated common data infrastructure challenges and developed short- and medium-term strategies to strengthen equitable pandemic response efforts. The following case studies do not tell the story of the full range of government agency services and activities but, rather, were chosen to offer a window into data-reliant COVID-19 response efforts and their evolution. The report also includes recommendations for federal leaders seeking to transform public health data infrastructure.



Definitions

Epidemiology: the study of how diseases develop and spread within a population.

Social determinants of health: the conditions in the places where people live, learn, work and play that affect a wide range of health risks and outcomes. These conditions can be grouped into economic stability, education access and quality, health care access and quality, neighborhood and built environment, and social and community context.¹

Health equity: conditions in which everyone has a fair and just opportunity to be as healthy as possible. Inequities are reflected in different health outcomes for different population groups and must be addressed by removing systemic obstacles to health.²⁻³

Public health surveillance: the ongoing, systematic collection, analysis and interpretation of health-related data essential to planning, implementation and evaluation of public health practices.⁴

1 <https://www.cdc.gov/socialdeterminants/about.html>
2 <https://healthequityguide.org/about/defining-health-equity/>
3 <https://www.cdc.gov/chronicdisease/healthequity/index.htm>
4 <https://www.cdc.gov/training/publichealth101/surveillance.html>

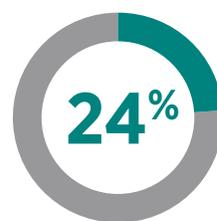


How Public Health Teams Adapted to Respond to the Crisis

Health Resources and Services Administration

Cultivating a flexible and adaptable workforce enables leaders to leverage existing employees quickly during urgent times.

The [Health Resources and Services Administration](#) within the federal Department of Health and Human Services provides grants and oversees nearly 1,400 health centers operating more than 13,500 delivery sites across the country as part of the [Health Center Program](#). These HRSA-funded health centers provide high quality, culturally competent, patient-centered health services to one in 11 people in the United States. Health center patients consist of particularly [vulnerable populations](#) including 62% who are people of color, 1.3 million people who do not have housing, 5.2 million people who live in or near public housing, nearly 980,000 migrant or seasonal agricultural workers, and 24% who are individuals best served in a language other than English.



Percentage of Health Center Patients best served in a language other than English.

Each year, HRSA receives data from organizations participating in the Health Center Program on patient characteristics, services provided, clinical processes and health outcomes, patients' use of services, staffing, costs and revenues as part of a standardized reporting system to monitor health centers' operational capacity and identify opportunities for program improvement. When COVID-19 hit however, HRSA's Bureau of Primary Health Care had to quickly figure out how best to collect data from Health Center Program participants to enhance public health surveillance—something the existing data collection strategy was not designed to do. As the bureau surveyed Health Center Program participants on a weekly basis to assess their operational capacity for telehealth implementation, COVID-19 testing,

prevalence and vaccination rates, it became clear that a significant amount of work was needed to modernize the data infrastructure required to collect, validate and verify survey data to understand what was happening on the ground in order to effectively respond.

This challenge was further complicated when the bureau needed to gather and share data to better collaborate with federal agencies such as CDC, the National Institutes of Health, and the departments of Health and Human Services and Housing and Urban Development on clinical trials for vaccines, antibody research and vaccine distribution. For instance, the four agencies needed to identify health centers in emerging COVID-19 hotspots and match them across disparate data systems for vaccine orders, outreach and more. HRSA teams had to match addresses and geocode to link health centers, which had direct effects on resource distribution for COVID-19 testing, representation of people of color and vulnerable communities in clinical trials, and distribution of COVID-19 vaccines and treatment.

“Syncing timely data across disparate data systems was probably the largest challenge or barrier to conducting public health surveillance.”

—HANK HOANG, DEPUTY DIRECTOR OF DATA AND EVALUATION AT HEALTH RESOURCES AND SERVICES ADMINISTRATION

The existing silos and inconsistencies warranted a need for robust collaboration and new systems for the public health emergency response. HRSA also needed new processes and more granular data as well as an increase in workforce capacity and data fluency among employees to strengthen surveillance efforts.

After devising a COVID-19 data collection strategy, the bureau’s senior staff recruited a small group of specialized bureau staff to serve on a part-time and full-time detail to manage the influx of COVID-19 survey data from health centers. As a result of this quick transformation, data analysts and knowledge management experts were able to efficiently harmonize large amounts of data across health centers, enhancing HRSA’s ability to distribute resources such as testing supplies, personal protective equipment and treatments for COVID-19 to underserved communities. In addition to sustaining continuity of operations for mission critical work of the bureau, the group, which grew over time, also worked to develop automated data flows, data validation processes and report generation and distribution systems. These processes enabled teams to analyze weekly survey data that was used to inform COVID-related response decisions across the entire agency.

“Data fluency is so powerful. You can have all the data in your hand but if your staff and your workforce aren’t comfortable with navigating, utilizing or analyzing it, it’s just going to sit there.”

–HANK HOANG, DEPUTY DIRECTOR OF DATA AND EVALUATION AT HRSA

Eventually, HRSA and CDC launched the [HRSA Health Center COVID-19 Vaccine Program](#) to help distribute vaccines throughout the country, informed by the surveillance and data collection strategy. To do this equitably and effectively, the bureau set up points of contact at multiple federal agencies to regularly exchange and validate critical data, such as up-to-date jurisdictional data for correctly identifying the most vulnerable communities. Collaboration with CDC was particularly critical in identifying health centers interested in participating in the Health Center COVID-19 Vaccine Program and linking them to the vaccine ordering systems. Further, with ongoing quality-improvement measures in place, many data scientists and statisticians across departments and agencies who did not previously communicate with one another established frequent check-ins to work through different data definitions, calculations or other data incongruities. By September 2021, the program had administered roughly 6.4 million vaccines, 75% of which were for people from communities of color.

BEST PRACTICES FOR FEDERAL LEADERS

- Quickly coordinate among organization leaders to outline emerging needs and priorities to realize the potential of a diverse and agile workforce.
- Identify points of contacts on data teams within and across agencies and establish frequent check-ins while delivering programs to coordinate on data sharing practices and standards and better meet mission outcomes.



Office of Innovation, New Jersey

Embracing innovation in data software and technologies during times of crises can help strengthen response efforts—even if it means starting over.

In spring of 2020 at the start of the pandemic, New York City was the [biggest hotspot](#) for COVID-19 cases in the country for communities of color. Bordering New York City, New Jersey faced the enormous risk of having the virus spread uncontrollably to its 9.4 million residents. To prevent this, the state [Office of Innovation](#) partnered closely with the New Jersey Department of Health and several other New Jersey agencies to support more than 560 localities and 99 local health departments monitoring the spread of the virus.

Mike Flowers had just started a new position as a visiting senior innovation fellow in the office when he was asked—within 48 hours of starting his new role—to redirect his focus to support New Jersey’s contact-tracing efforts. He and other state agency leaders quickly noted a fundamental challenge: Each local jurisdiction had authority over its own public health data collection, which meant there was no standardized approach across the 99 local departments.⁵ “The biggest challenge was the heterogenous data collection happening across the state. Different definitions, taxonomies and ontologies made comparing apples to apples from North Jersey to South Jersey nearly impossible,” Flowers said. While assisting local offices with sharing data, his team noted other barriers to navigating multiple levels of bureaucracy, as state regulations and procedures around disease reporting data were not designed to handle the volume of data in a pandemic.

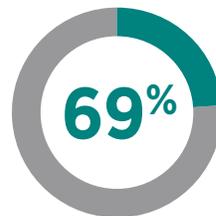
To develop a solution to this decentralized data infrastructure, state agency leaders knew they needed a team of epidemiologists and public health experts front and center. Through formal and informal outreach, they convened health experts within the agencies—many of whom underscored the need for a single contact-tracing platform to increase data standardization. Based on this input, the New Jersey governor issued an [executive order](#) on May 6, 2020, that mandated statewide use of a single contact-tracing platform to streamline processes and create consistent data definitions.

To decide on a platform, cross-agency teams identified software requirements by prioritizing the needs of epidemiologists and public health experts. These health experts helped identify

⁵ According to the United States constitution, public health surveillance systems are established from the states’ police powers, often passed on to local health departments, and allow them to regulate and restrict private interests for the public good. www.cdc.gov/eis/field-epi-manual/chapters/Legal.html

the need for a cloud-based platform that would facilitate consistency in data collection from over 1,000 contact tracers across the state so they could conduct virus surveillance—otherwise the data being reported on spread would be meaningless. Given the evolving nature of the pandemic and data demands, the experts also requested an agile platform that could have various software features added as data needs evolved. After the state team spoke with multiple vendors and examined what neighboring states were using, it identified a software vendor that met various technical demands and purchased licenses for 21 counties. This required the 99 local health offices to work together and coordinate on data collection so the platform could synchronize data from all local jurisdictions and streamline incoming data requests from the state.

State offices acknowledged the risk of upsetting local health officers while changing software systems in the middle of the pandemic. To garner trust and support around the new platform, the New Jersey Department of Health worked with the software vendor to host office hours and provided daily technical and scenario-based training to local officials and contact tracers across the state. Through these discussions, the benefits of standardizing data across the state quickly became apparent among local departments. These office hours laid the organizational foundation for local and state actors to collaborate and iterate on the software to ensure it helped contact tracers and managers collect consistent and accurate data and conduct surveillance. For instance, teams added a feature that enabled geotagging of cases that could be assigned to local contact tracers to help trace cases in their own community. As of [July 2021](#), a team of 1,700 contact tracers reached 69% of people infected by the delta variant within a week of using the single contract-tracing platform. If all states quickly share this level of data with national partners like CDC, it would be easier for leaders to create appropriate policies and distribute resources.



As of [July 2021](#), contact tracers reached 69% of people infected by the delta variant within a week.

BEST PRACTICES FOR FEDERAL LEADERS

- Invest in agile technologies and workforces that are prepared to accommodate changing needs in emergencies.
- Put public health experts and epidemiologists front and center when responding to a public health crisis.
- Exercise strong risk management and communication when implementing innovative strategies.



National Center for Advancing Translational Sciences

The federal government can be a leader in harmonizing data across sectors to drive data sharing, cross-sector collaboration and critical research.

The [National Center for Advancing Translational Science](#) within NIH supports the timely development of treatments and cures for diseases to improve individual and community health. Through laboratory, clinical and community research, the center's teams bring research communities together to study diseases so they can design and test interventions—which enables them to efficiently create more treatments to more patients. From rare disease research to treatment development for opioid addiction, the center's areas of encompass a wide range of health topics.

Translating research into clinical advances is often met with insufficient resources and talent and inefficient processes. For instance, one of the types of data most valuable to the center's work is [electronic health records](#)—a comprehensive record of a patient's medical history containing administrative, clinical, laboratory and diagnostic data.⁶ These records can be used to better understand how diseases affect diverse patient populations, inform the design of clinical studies and trials, and identify effective treatment interventions and care practices. However, because hospital systems often use [different electronic systems](#) with varying data formats and definitions, comparing records across patient populations has proved difficult.

The pandemic amplified the need for federal, state and private institutions to have a secure and timely data-sharing platform with important medical records. With funding from the CARES Act, a team of scientists at the center came together in April 2020 to create the [National COVID Cohort Collaborative](#), a robust, centralized data repository that could collect and harmonize electronic health records from multiple sources. The resource would also have built-in data analytic features to spur innovation in research.

Creating large health data-sharing platforms poses huge challenges due to laws designed to protect patient confidentiality, competition between institutions, and institutional trust. To address these concerns among the data-sharing community, scientists from the center worked closely with university and industry partners to develop a number of data security and protection measures.

⁶ In 2009 the Health Information Technology for Economical and Clinical Health Act provided as much as \$36.5 billion dollars of congressional funding to help health care providers across the United States use electronic health records for patients. See: www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf

“The social engineering was as difficult as the technical engineering. When developing governance for data access, we had to be mindful of maintaining institutional trust, addressing privacy concerns and emphasizing collaboration over competition.”

–DR. KENNETH GERSING, INFORMATICS DIRECTOR AT THE NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES

Three levels of protected data were offered: limited data sets that exclude selected personal identifiable information; data sets that remove personal identifiers and truncate or algorithmically shift identifying information; and synthetic data sets that resemble patient information statistically but are computationally derived, as another approach to avoid working directly with sensitive or private data. The collaborative also developed numerous [data transfer and data-use agreements](#) to outline exactly how data would be shared and used between partners. For instance, NIH guaranteed partners that they would receive free access to the data through a secure cloud-based platform. To inspire collaboration and scholarship over competition, organizers required participating researchers to set up a free [special ID](#) when registering for the portal that would track their future research and publications using data from the collaborative to properly recognize their innovative academic contributions.

Once partners share their data in their own preferred data format with the collaborative, a data analytics team runs quality checks and harmonizes the data using one common data model that converts electronic health records with different data definitions into one standard language. With this foundation, the collaborative grants researchers and physicians access to larger, more reliable data sets than they would have otherwise.

In September 2020, the team officially and successfully launched the data repository. By [October 2021](#), the resource features almost 3 million COVID-19-positive patients and more than 9 billion rows of data—one of the largest COVID-19 patient data sets in the world. Dr. Gersing estimated that approximately 1,300 investigators from 87 universities, hospitals, federally qualified health centers and other organizations across the country worked together to share information on test results, medications, procedures, medical conditions, demographics and more.

The patient-level clinical data is [updated](#) about twice a week and has built-in data analytic tools available to its users, including machine learning and predictive modeling. These features allow for novel analyses and stronger predictions when research teams are trying to study the effectiveness of a treatment.

Dr. Gersing said the breadth and dexterity of the collaborative has empowered researchers outside of government to examine a range of topics that may otherwise have been impossible to study. From assessing racial inequality in COVID-19 testing to understanding how symptoms present among at-risk groups, several studies have sought to [advance health equity](#) among vulnerable patient populations affected by the pandemic. As of August 2021, 248 projects and 223 data-use agreements have been issued by the collaborative.

248 projects
223 data-use agreements

Number of N3C projects and data-use agreements.

“The only way we were able to create N3C [the collaborative] was to say: ‘the gaps in data aren’t going to be filled in by another academic institution or a private company for that matter. It will be in the federal government’s enclave, period.’”

–DR. KENNETH GERSING, INFORMATICS DIRECTOR AT THE NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES

BEST PRACTICES FOR FEDERAL LEADERS

- Provide transparency through data use and data transfer agreements to build institutional trust when sharing data between partners.
- Consider developing a way to recognize work done by partners to encourage collaboration with data.



California Health and Human Services Agency

Having a shared vision focused on people, not programs, can help facilitate health data sharing and governance initiatives, strengthening government’s ability to serve the public.

The [California Health and Human Services Agency](#) is the largest agency in the state’s executive branch, comprising 12 departments, five offices and 33,000 employees. The agency provides a wide range of health, prevention and health care services—including 200 public health prevention programs alone—and was struggling to harmonize approaches to data analytics and disease surveillance systems across departments and offices.

Scott Christman, formerly the chief data officer at the state’s Department of Public Health, noted that 12 different health and human services departments were essentially doing business 12 different ways, specifically around data collection, analyses and reporting. He and other agency leaders recognized an opportunity to transform data operations to be more coordinated and streamlined, making the production of and access to data more efficient for staff and end-users alike. “We thought, wouldn’t it be more meaningful to the people of California whom we serve, if each of the 12 health departments did things the same way using data? This required us to think about an entire culture shift at the agency [when it came to addressing data operations],” Christman said.

12 departments
5 offices
33,000 employees

Organizational composition of the California Health and Human Services Agency.

With the goal of improving publicly offered health services for the people of California, agency leaders connected with the 12 health departments and the communities they serve to better understand how they use and benefit from agency and department data. Also inspired by the state of New York, leaders began championing the idea of having the health agency commit to data transparency, starting with data that was already publicly available online as PDFs and spreadsheets. By targeting early adopters among the health departments, in 2014, organizers convinced the agency to move towards data transparency and adopt an open data program. The statewide [Open Data Program](#) was designed to [increase access and sharing](#) of all publicly available data on a single platform.

Over the next two years, health department directors established a subcommittee under the new data program to help align teams and programs on new data policies and procedures and embrace a series of innovation initiatives focused on data operations. The group met monthly to develop basic shared governance for the platform, prioritizing what data should be made public first. By 2016, the group published data sets from all 12 agency departments using agreed-upon open data standards. The health agency also negotiated a single data-use agreement to share health agency data sets across the departments and with other agreed-upon internal partners, which included defining ways to securely and effectively share sensitive data that could include personally identifiable information. According to Christman, this internal commitment to data sharing drove a huge culture shift within the departments and led to wider recognition of data and the sharing processes as a critical asset among staff, partners and the Californians they served.

“Whether providing food stamps or Medicaid benefits, [health] agency departments all serve the same people of California. Shifting from being program-centered to person-centered helped multiple departments coalesce around the idea of data sharing during the pandemic.”

—SCOTT CHRISTMAN, FORMER CALIFORNIA DEPARTMENT OF PUBLIC HEALTH CHIEF DATA OFFICER

By the time COVID-19 emerged, existing data-sharing practices made data more readily available and easier to exchange with external stakeholders. Without the need to renegotiate new, one-off data agreements or build specific platforms to share data, the single-platform open data portal democratized information for the media, hospitals, community groups and the general public. Further, the internal data sharing agreement laid the foundation for a COVID-19 vaccine data-use agreement with CDC, helping California respond efficiently.

The state’s open data platform also enhanced its ability to respond to health equity challenges during the pandemic by upholding a commitment to transparency. Teams built numerous data dashboards with health equity measures—such as covid19ca.gov—to bring attention to health disparity concerns and help distribute vaccines to the most vulnerable and impacted areas. Data from these dashboards was shared on the open data portal which, in turn, was used by the media and other stakeholders—helping to ensure consistency in data reporting throughout California.

The disparities in our diverse communities are severe

COVID-19 disproportionately affects California's low income, Latino, Black, and Pacific Islander communities, as well as essential workers such as those in health care, grocery, and cleaning services.



Note: This data is cumulative since the first COVID-19 case was reported in January 2020. Case rate is defined as cumulative COVID-19 cases per 100K population. Death rate is defined as cumulative COVID-19 deaths per 100K.

The covid19.ca.gov/equity dashboard was California's first commitment to visualizing health disparities at the state level. It is fully accessible, translated in multiple languages and works on low-bandwidth phones.

BEST PRACTICES FOR FEDERAL LEADERS

- Work with enthusiastic early adopters to champion a new data initiative.
- When creating new data-sharing policies, focus on a single data exchange agreement between internal partners for greater efficiency during emergency response.
- Ensure cross-sector representation in agency task forces to develop informed data-sharing policies and governance structures.
- Prioritize end-user needs when developing data governance structures, policies and procedures.
- Prioritize inclusivity when building and designing data tools, resources or visualizations to make them fully accessible to users.



Center for Preparedness and Response, Division of Emergency Operations: CDC

Proactively investing in governance efforts and robust planning, especially with partners, can equip agencies to respond rapidly and effectively during a crisis.

Staffed 24 hours a day, seven days a week, 365 days a year, the [Emergency Operations Center](#) at CDC is equipped with highly trained experts who monitor and respond to public health threats. Some of these experts work on the Situational Awareness team in the Center for Preparedness and Response. The center works with agency partners to leverage advanced technology and processes to analyze critical data and information and develop tools leaders can use to help keep people safe during public health emergencies.

During a public health emergency, the team curates data on cases, deaths, lab results, hospitalizations, emergency department visits, vaccinations and social vulnerability to analyze demographic trends and generate daily reports, dashboards and briefings for leadership. These reports provide timely and accurate information on disease severity, affected populations and countermeasures for creating policies and allocating resources to mitigate disease impact.

For the Situational Awareness team, planning is key for emergency operations. For instance, the team proactively develops necessary relationships with international, federal, state, local, academic and nongovernmental partners to prepare data and set up data agreements to collaborate more efficiently during an actual emergency. “Seventy percent of what we do should focus on preparedness, meaning preparing our data and getting agreements established with partners should all be done ahead of time so that when an event happens, teams are not struggling and scrambling—because there is a plan in place,” said Jim Tyson, the chief of the Situational Awareness team at CDC.

When the COVID-19 pandemic hit, however, the team faced unprecedented barriers to its work. Early on, for example, states were focused on the demands of their own response efforts. That sometimes meant there was limited time and resources to devote to managing real-time data sharing with federal agencies. The data the Situational Awareness team needed to produce decision-support products that CDC leadership and agency partners depend on were often incomplete and in a format that required significant manipulation before it could be analyzed, or simply unavailable.

One way the Situation Awareness team navigated these challenges was by using “web-scraping” to collect COVID-19 cases and deaths directly from jurisdictional partner and school-district websites. Although this began manually, the team partnered with CDC COVID-19 leadership and Johns Hopkins University’s Applied Physics Lab to automate this process and aggregate case counts to track the spread of the disease within states and across the country in a timelier manner. To ensure the data was accurate, the team coordinated daily with nongovernmental partners, including Johns Hopkins University’s Coronavirus Resource Center, to identify anomalies in data reporting and validate one another’s data. Tyson believes these exchanges were only made possible because every partner recognized that they could not conduct public health surveillance alone. Because this was the first time

“The key here was excellent coordination and collaboration in near real-time through cross-jurisdictional mutual support efforts. It wasn’t so much technology. It was people, culture and process.”

–JIM TYSON, BRANCH CHIEF, SITUATIONAL AWARENESS BRANCH AT THE EMERGENCY OPERATIONS CENTER, CENTERS FOR DISEASE CONTROL AND PREVENTION

SARS-CoV-2—the virus that causes COVID-19—had been reported, some data elements, value sets and standardized codes were not in place to report this new and novel virus. To address some of the data standardization issues, the team coordinated with CDC’s COVID-19 response leadership, the Center for Surveillance, Epidemiology, and Laboratory Services, and other CDC centers and partners to create a CDC COVID-19 Information Management Repository. The repository provided states and public health partners a trusted source for COVID-19 informatics resources, including data requirements for COVID-19 data reporting. The COVID-19 Information Management Repository leveraged data interoperability preparedness work over the years between the Situational Awareness team, the Department of Health and Human Services, the National Health Coordinator, and CDC’s Public Health Information Network Vocabulary Access and Distribution System. This repository is maintained by CDC to help improve interoperability of surveillance data systems among selected state, territorial, tribal and local health care and public health partners.

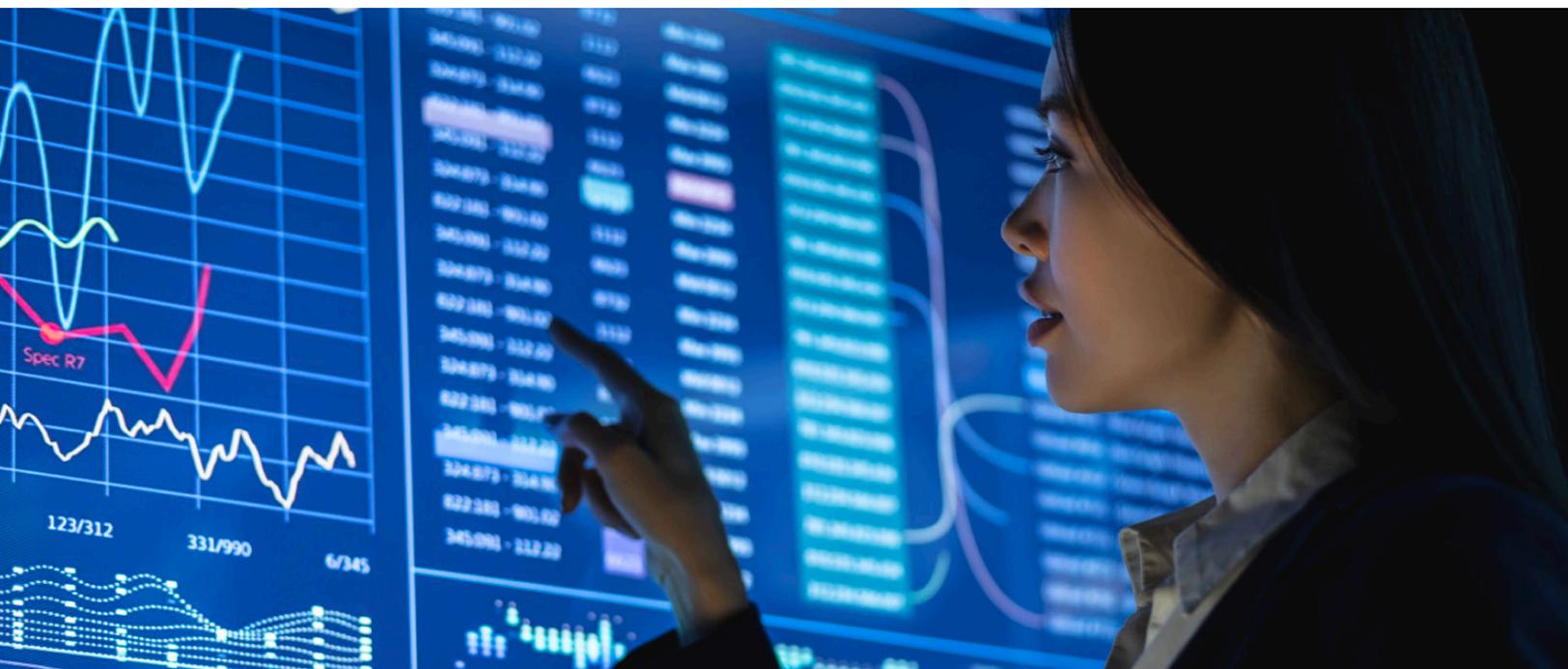
The team was resilient to these data challenges, and noted critical lessons learned, particularly around transforming planning processes for all types of emergencies. “There needs to be a long-term strategy during preparedness activities to establish stronger enterprise systems and data standardization for availability, access and sharing to strengthen future operational responses,” said Roger Harlan, knowledge management team lead at the Situational Awareness team. Organizations have an opportunity to make better use of emergency planning tools that agencies such as the Department of Defense and the Federal Emergency Management Agency use effectively, known as “information exchange requirements” and “information collection planning,” he added. These tools help strengthen processes and governance by requiring agency teams to talk through issues related to plans, roles, partners, data capacity and associated risks prior to any emergency.

WORKFORCE TIP

When the team faced high turnover and burnout among skilled employees who had experience in emergency situations, the management team worked with other agencies and universities to bring in additional personnel for a two- to six-month rotation and train these individuals on COVID-19 response operations. Many of these temporary employees learned quickly how to produce timely data reports, gaining new software skills through screen-sharing—for example, learning Power BI, a data visualization tool. These employees helped the team build short-term capacity and minimize burnout. To learn more about talent exchanges across the federal government, please see the Partnership’s [report Trading Places](#).

BEST PRACTICES FOR FEDERAL LEADERS

- Work with state partner organizations when developing data definitions and standards for states—for example, the Council of State and Territorial Epidemiologists, the Association of State and Territorial Health Officials, and the Public Health Informatics Institute.
- Recruit staff from other agencies or universities on a rotational basis to fill short-term workforce needs.



Johns Hopkins Coronavirus Resource Center

Data scientists can help decisions-makers become more equitable by collecting, visualizing and publicly sharing accurate and comprehensive data.

In March 2020, Johns Hopkins University launched the [Coronavirus Resource Center](#)—a robust COVID-19 data-tracking website that garnered global attention by providing the public with critical information on coronavirus. What started as a team of 20 data scientists manually collecting information from more than 3,000 state and county health departments expanded to become a prominent global resource with [more than 1 billion visits](#) as of January 2021.

As the resource center attempted to collect COVID-19 data on testing, confirmed cases, hospital bed capacity, morbidity and more, it quickly found that local, state and federal government did not have [standard guidance](#) for data collection. “The most important element of a strong data practice is actually in the governance and alignment of creating a common language and rules around how and why data is collected and applied to problem solving,” [said Beth Blauer](#), executive director of the Centers for Civic Impact and data lead of the resource center. She added, publishing COVID-19 data by race at the state level helps highlight how racial health disparities and social determinants of health affect how the virus spreads throughout the country.

1,000,000,000
site visits as of January 2021

Website traffic as of January 2021.

To fill these critical gaps quickly, Blauer said the resource center proactively and regularly consulted with multidisciplinary professionals such as epidemiologists, medical doctors and public health researchers to ensure the website was based on the highest standards of science. The data science team met frequently to discuss data structures and standard definitions for common data terms such as demographics, tests, deaths and cases, to achieve consistency across all levels of government. Once the data structures and definitions were vetted with experts, the team developed scripts to automate data collection from state websites and other reliable sources of information to harmonize and publish this data.

The resource center used [sophisticated maps and visuals](#) to empower leaders to respond with evidence-based decisions, using available information. These data visualizations became an important tool for health equity and accountability. When the team created national maps for COVID-19 testing, case and death counts, which contained breakdowns by race and ethnicity, it was clear that a handful of states were not collecting this important demographic data and appeared as outliers as a result. The highly visible website created an incentive for these states to submit missing demographic data, which provided a more complete and accurate representation of COVID-19’s impact on all communities. Without this [data](#), policymakers and other decision-makers would not be able to effectively target the most vulnerable communities.

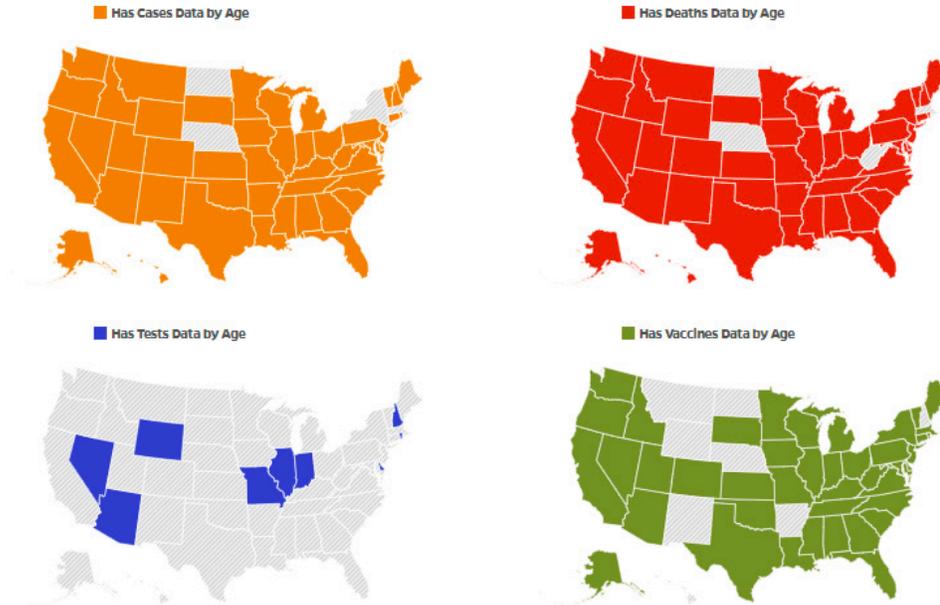
“Inconsistencies in categorization between states and even within states make [demographic] data not comparable and can obfuscate the disproportionate effects that the pandemic—and, in reality, across all programs targeted at entrenched social determinants of health—have had on people of color.”

—BETH BLAUER, EXECUTIVE DIRECTOR OF THE CENTERS FOR CIVIC IMPACT AND DATA LEAD OF THE CORONAVIRUS RESOURCE CENTER

Age Ethnicity Gender/Sex Race Race & Ethnicity

SELECT A GROUP

All



STATE	CASES	DEATHS	TESTS	VACCINES
Alabama >	✓	✓	✗	✓
Alaska >	✓	✓	✗	✓
Arizona >	✓	✓	✓	✓

The coronavirus.jhu.edu/data/racial-data-transparency website visualizes demographic tracking of demographic data related to COVID-19. It became a powerful incentive for states to collect and submit demographic data.

The team continued to build its workforce capacity by staying up to date on the most important data science innovations and attending technical skills-based training sessions hosted by federal agencies. The team standardized data definitions, automated data collection and worked closely with state and local partners. By using these best practices to collect, analyze and visualize data, the website has helped the public, policymakers and health care professionals respond to the pandemic.

BEST PRACTICES FOR FEDERAL LEADERS

- Organize crucial skills-based training for cross-sector workforces.
- Provide national data standards to improve data collection processes and strengthen equitable mission outcomes, especially for health equity-related metrics such as how race and ethnicity should be defined and collected.
- Partner with local and state agencies to establish standards, create automated data collection processes and solve problems.



Recommendations

While public health threats remain imminent, it is critical for congressional and federal leaders to make substantial and timely progress on modernizing the nation’s public health data infrastructure to improve timeliness, quality, coordination, burden-sharing and technology integration of public health data. The following recommendations are for leaders in Congress and federal agencies at the forefront of public health response, such as CDC, the National Institutes of Health and other divisions in the Department of Health and Human Services. The Chief Data Officer Council can play a coordinating role in setting standards, motivating talent and generating guidelines and best practices for cross-government agreements. These recommendations were informed by several public health organizations, experts and advocates as well as prominent policymakers at the state and federal levels.

Invest in core public health surveillance systems to be equipped for all public health threats. Federal government funding for public health threats has historically been tied to specific diseases, resulting in fragmented, siloed systems and processes, which sometimes extend to the state level. To bring transformation to data infrastructure, Congress must help federal agencies such as CDC invest in sustainable and predictable funding, infrastructure policies, systems and technologies that can be iterated and built upon—including cloud computing, artificial intelligence and machine learning. A system that is designed to integrate a broader range of major public health datasets, including immunization registries, will better equip cross-governmental teams with the tools needed to respond to all future public health crises.



Advance health equity outcomes by providing clear data standards and definitions and addressing systemic disparities. Data quality and governance were widely cited as some of the most common barriers to navigating data challenges during the pandemic, especially when trying to serve historically marginalized populations. To improve data quality and fill in critical missing information, federal agency leaders should develop and provide standardized data definitions and reporting guidance, particularly for health equity and demographic data. Though agency roles in such efforts may vary, the Chief Data Officer Council is well-positioned to translate its COVID-19 public health data coordination efforts into lessons, guidance, definitions and requirements for agency partners. Through cooperative agreements that focus on advancing health outcomes, the federal government can require state and local actors to adhere to these guidelines to qualify for federal funding. Leveraging data visualizations and tools can also help provide accountability and strengthen government’s ability to target interventions to



the most vulnerable communities. To advance progress in healthy equity, federal leaders responding to public health crises should use social determinants of health frameworks to examine broader structural drivers of health.

Implement policies and strategies to make public data more accessible, timely and secure.

Several experts underscored the need for the federal government to share data in its most raw form as early as possible to enable cross-sector organizations to efficiently use, adapt and analyze that data according to their own needs. To facilitate this type of data sharing, the administration and congressional leaders should implement policies and legislation to incentivize data sharing and address important privacy concerns, such as for patient-generated data.⁷ These leaders may want to explore how to resolve patient matching. They may also want to require data-use agreements to regulate clear purposes for how data will be used by each partner involved.



Invest in a more robust, flexible and agile federal workforce. A skilled workforce is critical to managing and implementing a strong data infrastructure as well as ensuring the data itself is sound enough to use and analyze. The range of talent necessary for such work is wide-ranging. Data scientists and engineers are as important as leaders, program managers, administrative staff and communicators with fluency in requesting, understanding and communicating data. With high turnover, burnout and limited capacity, federal public health teams worked quickly to deploy, recruit, restructure and transition their workforces during COVID-19. Some of this included conducting remote training, partnering with universities to attract short-term talent or standing up temporary offices and positions to address data needs. Federal agency leaders should continue to document lessons learned from these strategies that helped agencies prioritize pandemic response efforts and determine what may be replicable and scalable. They should also continue to [invest in and retain skilled employees](#) who can work with data and advanced technologies.



Put public health experts front and center on public health response decisions. When responding to a public health crisis, federal and state leaders should invest in agile data technologies, systems and processes that best support public health professionals—even if it means starting fresh. To make well-informed decisions, policymakers and leaders should consult with epidemiologists and other health-focused data scientists about what and how data should be defined, collected and reported.



Strengthen cross-governmental collaboration around public health data. The federal government should provide more support to state- and county-level governments in future public health crises through clear reporting guidance, efficient funding protocols and workforce development. Some strategies may include providing workforce skills-based training virtually or streamlining report requirements to reduce the burden on state and local offices. Because states generally have technical and regulatory control over public health data collection, it is crucial that the federal government partners with relevant state and local health departments and public health organizations that can help inform, connect and advise the federal government on state needs. Some entities may include: the Council of State and Territorial Epidemiologists, the Association of State and Territorial Health Officials, the National Association of County and City Health Officials, the Public Health Informatics Institute and the American Immunization Registry Association.



⁷ CODE has a list of additional strategies here: healthdatasharing.org/wp-content/uploads/2020/08/PCORI-RT-Series-Summary-FINAL-2020.07.30-web.pdf

“The most powerful preventative measure in a public health crisis is information.”

—RYAN PANCHADSARAM, CO-FOUNDER, COVID EXIT STRATEGY

Conclusion

Strengthening cross-sector and cross-governmental public health data collaboration, collection and use must be a priority of the current administration as well as agency leaders. Partnerships between the federal government and state, local and health care partners have recently developed out of need, due to COVID-19. These partnerships should be further developed to create long-term data use agreements and sharing. In line with the Open, Public, Electronic and Necessary Government Data Act, federal agencies should continue to publish all data possible in standardized formats. Leaders within the federal government such as the Federal Data Strategy Community of Practice and the federal Chief Data Officer Council should further develop guidelines and create actionable plans to help agencies operationalize overarching guidance.

The CARES Act provided CDC with \$500 million for the Data Modernization Initiative to improve public health surveillance. That investment should contribute to cross-sector collaboration, standardizing data definitions, investing in workforce trainings, transparent decision-making, data collection and aggregation processes to further improve the accuracy, quality and timeliness of data.

Project Team

Xiaowen Cui

Intern, Research, Analysis and Evaluation

Loren DeJonge Schulman

Vice President, Research, Analysis and Evaluation

Samantha Donaldson

Vice President, Communications

Cara Glancey

Events Associate Manager

Mikayla Hyman

Associate, Research, Analysis and Evaluation

Claire Mills

Intern, Research, Analysis and Evaluation

Ellen Perlman

Senior Writer and Editor

Audrey Pfund

Senior Design and Web Manager

Netanya Quino

Intern, Research Analysis and Evaluation

Reema Singh

Manager, Research, Analysis and Evaluation

Max Stier

President and CEO

Acknowledgements

The individuals listed below generously offered their input on this report. We greatly appreciate their time and counsel. However, the contents of this report do not necessarily reflect the views of those we interviewed. Additionally, the views of participating federal officials do not necessarily reflect positions or policies of the federal government or its agencies.

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Jacqueline Burkholder

Deputy Branch Chief, Situational Awareness Centers for Disease Control and Prevention

James Tyson

Branch Chief, Situational Awareness Centers for Disease Control and Prevention

Roger Harlan

Emergency Management Specialist, Knowledge Management, Situational Awareness Centers for Disease Control and Prevention

Stephen Soroka

LIMS Senior Scientific Advisor Centers for Disease Control and Prevention

Lesliann E. Helmus

Associate Director for Surveillance, Division of Health Informatics and Surveillance Centers for Disease Control and Prevention

Ileana Arias

Associate Deputy Director, Division of Public Health Science and Surveillance Centers for Disease Control and Prevention

Melvin Crum

Computer Scientist (Informatics) Centers for Disease Control and Prevention

Heather Strosnider

Senior Advisor for Surveillance and Data Modernization Centers for Disease Control and Prevention

Joel Cohen
Director, Center for Financing, Access and
Cost Trends
Agency for Healthcare Research and Quality

Hank Hoang
Deputy Director, Data and Evaluation Division
Bureau of Primary Health Care, Health Resources
and Services Administration

NATIONAL INSTITUTES OF HEALTH

Dr. Ken Gersing
Informatics Director, National Center for Advancing
Translational Sciences

COVID EXIT STRATEGY

Ryan Panchadsaram
Co-Founder

CALIFORNIA DEPARTMENT OF PUBLIC HEALTH

Scott Christman
Chief Deputy Director, Office of Statewide Health
Planning and Development

NEW JERSEY OFFICE OF INNOVATION

Mike Flowers
Senior Fellow

VERMONT DEPARTMENT OF HEALTH

Veronica Fialkowski
Health Surveillance Epidemiologist,
Division of Health Surveillance

Lauren Prinzing
Health Surveillance Epidemiologist, Division of
Environmental Health

AMERICAN IMMUNIZATION REGISTRY ASSOCIATION

Mary Beth Kurilo
Senior Director of Health Informatics

NATIONAL ASSOCIATION OF COUNTY AND CITY HEALTH OFFICIALS

Adriane Casalotti
Chief of Government and Public Affairs

THE ASSOCIATION OF STATE AND TERRITORIAL HEALTH OFFICIALS

J.T. Lane
Senior Vice President, Population Health & Innovation

JOHNS HOPKINS UNIVERSITY

Beth Blauer
Data Lead, Coronavirus Resource Center

CENTER FOR OPEN DATA ENTERPRISE

Joel Gurin
President and Founder



1100 New York Ave NW
Suite 200 East
Washington, DC 20005

ourpublicservice.org
(202) 775-9111

 [partnershipforpublicservice](https://www.facebook.com/partnershipforpublicservice)

 [@PublicService](https://twitter.com/PublicService)

 [rpublicservice](https://www.instagram.com/rpublicservice)

Cover Photo: shutterstock.com