CASE STUDY

How the Indiana Management Performance Hub's Data Capacity Helped Fuel COVID-19 Research and Response

AISP

Acknowledgments

"AISP Case Study: How the Indiana Management Performance Hub's Data Capacity Helped Fuel COVID-19 Research and Response" was created by Actionable Intelligence for Social Policy (AISP) at the University of Pennsylvania. It was developed through domain expert interviews, document review, and from many years working with and learning alongside Indiana MPH and other AISP Network sites.

We would like to thank Ashley Hungate, Joshua Martin, and Connor Norwood for their time and contributions to this case study. We'd also like to acknowledge the expertise shared as part of the 2021 MPH Data Day, which greatly informed this work. See more here.

MANAGEMENT PERFORMANCEHUB

Learn more about Indiana Management Performance Hub at their website: data.IN.gov

This case study was authored by Emily Berkowitz and Della Jenkins, with support from the whole team at AISP. It was made possible by grant funding from the Bill & Melinda Gates Foundation and the Annie E. Casey Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funders.

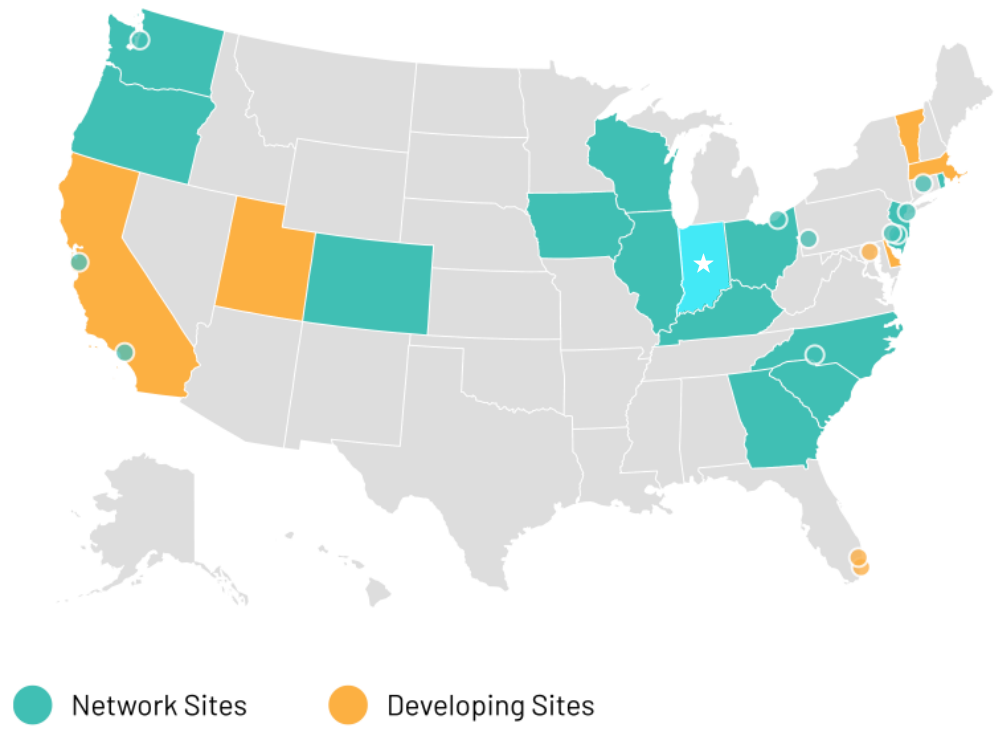
Suggested Citation

Berkowitz, E. & Jenkins, D. (2021). *AISP Case Study: How the Indiana Management Performance Hub's Data Capacity Helped Fuel COVID-19 Research and Response.* Actionable Intelligence for Social Policy. University of Pennsylvania.

Introduction

The economic and social impacts of the COVID-19 pandemic have heightened demand for cross-agency data capacity, as policymakers are forced to reconcile the need for expanded services with extreme fiscal constraints. In this context, integrated data systems (IDS) — also commonly referred to as data hubs, data collaboratives, or state longitudinal data systems – are a valuable resource for data-informed decision-making across agencies. IDS utilize standard governance processes and legal agreements to grant authority for routine, responsible use of linked data, and institutionalize roles across partners with shared priorities.

The following case study details how one state-level IDS, the Indiana Management Performance Hub (MPH), expanded existing capacity to facilitate groundbreaking COVID-19 research, support vaccine distribution, and inform long-term recovery efforts. While every context is unique, Indiana's experience offers lessons for other state (and local) governments seeking to build flexible capacity to integrate data for impact.



AISP Network Sites Map

Background

The Indiana MPH is an enterprise-level state data sharing effort that provides analytics and insights to improve outcomes for Indiana residents. MPH's primary charge is to act as a central data exchange where identifiers are encrypted and then linked for analysis. The office serves as a resource for partners throughout state government looking to leverage data for decision-making. MPH also provides external stakeholders with robust, centralized access to open data from 20 partner agencies.

MPH was created in 2014 through an executive order (EO) from the governor and implemented in collaboration with the state's Office of Management and Budget and Office of Technology. The EO gave MPH the authority to act as an agent of each agency they support and tasked them with central data stewardship for the state. In 2017, MPH's authority was codified into law (Chapter 26). Importantly, MPH received bipartisan support because of their emphasis on security: rather than relying on each agency to send, receive, and safeguard personally identifiable information, MPH became the office equipped with expertise and capacity to handle the sharing and linkage of sensitive data on behalf of agencies. Of course, stepping into this role required more than a mandate; it required that MPH build trust and demonstrate value with their agency partners. One way they did this early on was to facilitate a widescale effort to update reporting and performance indicators across agencies to make data more usable and accurate. Their work has paid off, and to date MPH analytics have guided state efforts to address a range of complex social policy challenges, including the opioid epidemic, infant mortality, recidivism and rearrest, and Medicaid optimization.

For more on the history of legislative and executive action on data sharing in Indiana and beyond, explore AISP's report, "Building and Sustaining State Data Integration Efforts: Legislation, Funding, and Strategies."

In March 2020, MPH also launched the Enhanced Research Environment (ERE) to allow additional external researchers to interact with Indiana data in a secure environment using prebuilt computing tools. The environment was designed to improve on the traditional research cycle, making it both more secure and more routine. Typically, researchers receive and download data to their personal computers and perform analysis using custom, locally hosted tools. With ERE, data are no longer downloaded but rather stay within the secure environment; there are standard tools for analysis, and results and research datasets can be easily shared. Now, only a year since it launched, ERE hosts over 100 users across 20 separate agencies and universities, and has 16 active projects.

COVID-19 Response Efforts

MPH's capacity, authority, and existing relationships left them well positioned to pivot rapidly to support state emergency information needs. MPH has played an essential role in COVID-19 prevalence and spread research, which was led by the Indiana Department of Health (DOH) and researchers at the Fairbanks School of Public Health at Indiana University–Purdue University Indianapolis (hereafter referred to as Fairbanks). At the start of the pandemic, Fairbanks and DOH set out to study disease spread and better understand the characteristics associated with worsened cases. However, DOH case data alone would not accurately measure the prevalence of COVID-19 because of asymptomatic cases and the fact that many people would not or did not present at a hospital or healthcare facility. This meant researchers would need a population dataset representative of the state, which only MPH could provide, as well as a new and coordinated data collection effort. As Chief Data Officer for Indiana Family Social Services Administration Connor Norwood succinctly put it during MPH Data Day 2021, "MPH created the denominator." MPH was an ideal partner for this task not only because of their technical expertise but also because their robust data governance and existing legal framework enabled the project to move forward quickly, without the need to renegotiate agreements.

Specifically, MPH supported the DOH and Fairbanks' effort by

- Constructing a population-level dataset to act as the denominator for this and other research studies, and
- Generating a list of contact information for people randomly identified for testing and study participation.

To ensure these data were actionable, MPH also worked with collaborators to

- Securely host the dataset for researchers and analysts to use for approved studies,
- Build and manage daily dashboards for the public to use in order to understand and track COVID-19 in their communities, and
- Pivot to support vaccination efforts.

In the following section, we describe each of these activities in more detail.

Creating the Dataset & Contact List

First, working with the Department of Revenue and Governor's Office, MPH researchers used tax records to create a 6.3-million-person file (including both adults and their dependents) to use as the denominator for a groundbreaking COVID-19 study. They also used this population data file to generate research-ready randomized control groups. Notably, MPH did not have access to these datasets prior to the pandemic, but their existing relationships with other public agencies as well as their clear governance processes and legal frameworks allowed this work to move forward quickly. They relied heavily on the domain expertise of the Revenue Department to support determinations about the most relevant and useful data elements for research purposes and to help move this process along.

Next, Fairbanks researchers used stratified random sampling across Indiana's public health preparedness districts to determine who would be contacted to participate in the study, which involved COVID-19 testing regardless of symptoms or exposure. MPH generated a list of contact information and provided it to researchers who took over outreach and subsequent steps in the study. MPH was able to cross-check tax filing data with Bureau of Motor Vehicles information to confirm contact information and reduce errors for the study sample. To uphold data privacy and security standards, MPH provided researchers with only the absolute fields needed for random sampling – such as age, rather than date of birth, and zip code, rather than full address. Following public announcement of the study, nearly 15.5 thousand Indiana residents were contacted to participate in testing in April 2020.

Results & Later Linkages

With novel testing data on asymptomatic people, researchers were then able identify the proportion of the sample who had COVID-19. Prior to the creation of the data file (and subsequent testing of randomly selected individuals from within it), understanding of infections was based solely on individuals with severe symptoms who sought medical treatment or who otherwise were tested for the virus. MPH's data linkages and the new prevalence study data collection effort changed everything. Preliminary analysis of the population-level data suggested 186,000 COVID-19 infections, while the state had only been aware of 17,000 cases through testing and hospitalization.

Initial findings also indicated a 44 percent asymptomatic rate, which was comparable to country-wide estimates found in Los Angeles, California. This percentage was the first of its kind and was widely cited by the Centers for Disease Control and Prevention (CDC) and public health agencies across the country. A second phase of testing was conducted in June 2020, allowing researchers to compare estimates and assess the effectiveness of prevention efforts (e.g., mask mandates, social distancing requirements). This information was crucial to the state's reopening plan.

The population dataset created by MPH was also used to report two important COVID-19 metrics — unique individual positivity rates and all positive results from all tests — to policymakers and the public. Unique individual positivity rates capture the rate of positive tests for each individual-level case of COVID-19, and require deduplicating multiple tests from the same person. This precise method provides a granular view of community spread and enabled the state to accurately determine how and where people were contracting the virus. Information about cases and positivity could then be linked to other health data – like information on preexisting health conditions, emergency room visits, and hospitalizations – to deepen understanding of risks associated with virus transmission and case severity. Separately, the "all tests" method calculates a rate of positivity based on the total number of tests administered, which can include multiple tests of the same individual. "All tests" is a more common approach, so the metric was crucial for "apples to apples" comparison across states.

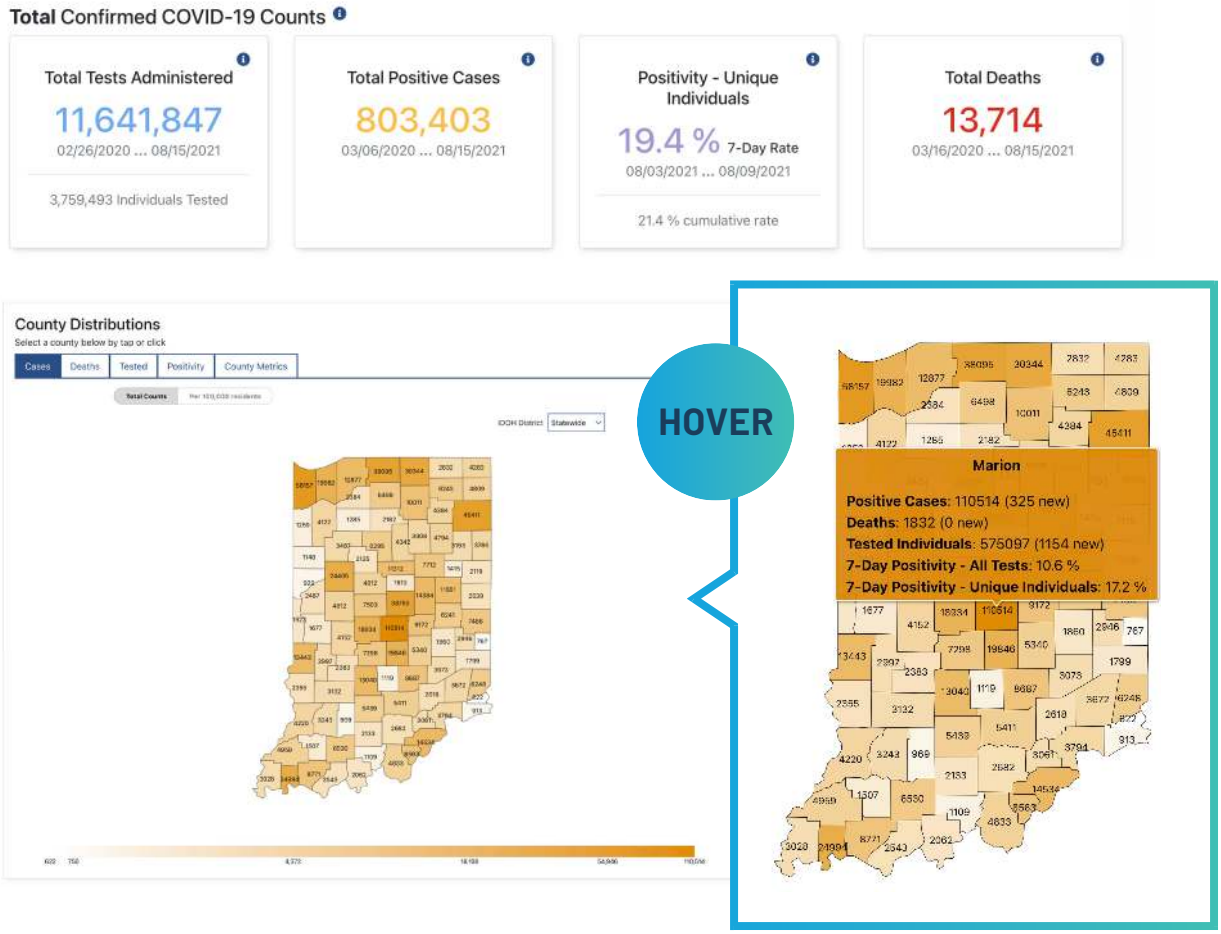
Making sense of these various measures and the risks associated with COVID-19 was a complicated endeavor and required additional waves of research and dialogue. Without research and data points about prevalence, transmission rates, and other key information to compare to, researchers had to make a lot of assumptions and manage many unknowns. For example, when modeling efforts first began, there was no consensus on the role of masks in preventing spread, and it was not yet understood how age or comorbidities contributed to mortality. Over time, assumptions could be minimized, and models were refined to reflect national trends, making them more reliable.

The prevalence research conducted by DOH, MPH, and Fairbanks had national implications: their findings have been used by the CDC to improve national public health policy response, as well as by journalists across the country to inform reporting and recommendations throughout the COVID-19 pandemic. Without these data and collaborative partnerships, COVID-19 response and recovery would have lagged substantially.

Providing Secure Access for Researchers & Dashboards for the Public

Once MPH had constructed the "denominator" dataset for the foundational COVID-19 prevalence study, they also realized they had a crucial role to play in providing secure access to these data for additional research. The full dataset is now hosted on the ERE in a granular format, and MPH's existing capacity enables timely approval of projects and execution of new research. A number of COVID-19 metrics, developed by the Family and Social Services Administration and DOH, are also available to researchers and analysts through the ERE.

In addition, data on a streamlined set of outcomes is presented in public-facing dashboards for the broader community, and updated daily. These data are anonymized before being made public. As the two figures below demonstrate, MPH and their partners present data in a way that is simple and accessible but provide options to click or hover for additional detail and definitions.



Reproduced from https://www.coronavirus.in.gov/ with permission.
A full list of Indiana's public-facing datasets related to COVID-19 is available here.

It is important to balance the public's expectations for reliable information with the necessary course corrections made to the models behind the numbers. Communication and transparency are key to finding this balance. The COVID-19 data response team made sure to own and explain any mistakes – what happened, how it happened, and the impact it had on data and information. A notable example occurred in December 2020 when Indiana experienced a short-lived software error that falsely increased positivity rates across the state's counties. During that same time, officials had also decided to adjust their methodology for calculating positivity rates as a result of increased testing. This combination of events could have led to major confusion and distrust, but State Health Commissioner Dr. Kristina Box moved quickly to provide a thorough explanation via a virtual news conference alongside other government officials. Information about the impact of the error as well as the impact of the planned changes in methodology was documented and made available to the public online. MPH personnel also opened up lines of communication to receive feedback and improve their work. Because of lockdown and quarantine measures, social media became an important tool for interacting with users: the agencies were able to use Twitter to engage with academics and make changes based

on their research needs. Facebook groups also became a space for crowdsourcing ideas and sharing findings using the publicly available data. Indiana's public data on COVID-19 received over 1.4 million views over the course of the last year, highlighting how much the community relied on the data infrastructure and information.

Pivoting to Vaccination Support

Now, in 2021, MPH is deeply involved in vaccine outreach and communication. When vaccines first become available, analysts worked to determine age brackets (based on date of birth in records) for targeted outreach. Cross-referencing with assisted living and nursing homes allowed them to deduplicate efforts to vaccinate older and higher-risk individuals. Contact was also tailored to the preferences of specific groups, such as sending texts to younger cohorts and postcards to older cohorts. To minimize risk and protect privacy, the minimum necessary amount of information about individuals was shared, and contact information like emails and phone numbers was kept as two separate files.

The process has paid off: before DOH began requesting information on targeted zip codes with low vaccine rates for mass vaccination events, they often found low uptake. In one instance, there was a 90 percent appointment vacancy rate in a high-risk neighborhood. MPH was tasked with providing a contact list of people within that zip code, and, within one hour of targeted communication, vaccination appointments were 100 percent full. MPH's rapid and effective data sharing has had lifesaving impacts and demonstrates the importance of investing in data infrastructure.

Lessons Learned

In many ways, the COVID-19 pandemic has demonstrated the value of building enterprise-level cross-sector data capacity. Because they had developed relationships and maintained trust, established their legal authority, and invested in secure infrastructure, MPH was able to support statewide COVID-19 response early and often, contributing to both local resource allocation decisions and nationally significant research findings on the spread and prevalence of this disease.

Below, we outline lessons learned from the Indiana approach that we believe may be applicable in other state and local contexts.

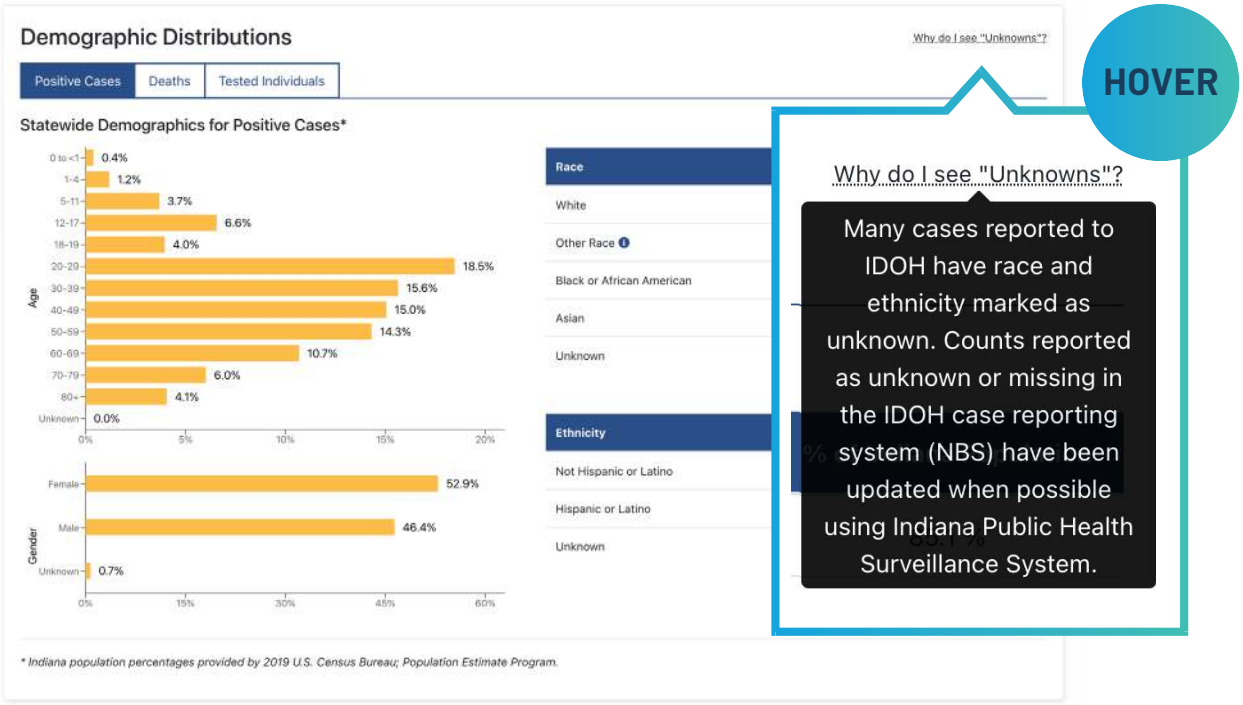
- **Invest in data infrastructure now.** Indiana prioritized cross-sector data capacity, and MPH laid the groundwork with strong governance, relationships, and technical tools long before the pandemic hit. These investments are paying off during COVID-19 and will continue to support future crisis response. For more on strategies for funding cross-agency data capacity, see AISP's report, "Building and Sustaining State Data Integration Efforts: Legislation, Funding, and Strategies."

> **"The work MPH has put into building out the State's data infrastructure has allowed us and our partners to mobilize quickly in response to COVID-19 and use data to drive our decisions in a meaningful way. The great thing is that this agile model can be used in multiple future scenarios, including guiding the State to its new normal."**
>
> —Josh Martin, Indiana Chief Data Officer and MPH Director

- **Prioritize security above all else.** Working across a large number of partners can make the legal and technical aspects of data integration more complex. In short, more data equals more risk. For that reason, MPH committed to protecting privacy and security at the beginning. The agency also frequently consults experts to ensure that they meet best practices across the legal, procedural, and physical components of data security. For more on a multidimensional approach to data security, see pages 17-18 of AISP's "Introduction to Data Sharing."

- **Learn in conversation with partners.** Knowing they alone did not hold all the answers, MPH relied on agencies to help interpret data on COVID-19. They also sought out new collaborations – with university partners and subject matter experts – in order to leverage new types of information for a public health crisis. MPH brought their capacity to the table but also recognized the strengths of their partners, working together to determine what data they had, what data they needed, and how best to transform datasets into actionable information for policymakers and the public.

- For example, Indiana Health Information Exchange staff helped the COVID-19 response team think through the importance of state demographics in how data were visualized and reported, despite limited-quality data on race and ethnicity. In Indiana, non-white and Hispanic populations had consistently higher positivity rates — a trend across much of the U.S. Policymakers, researchers, and subject matter experts have noted that a lack of reliable race and ethnicity data has been one impediment to developing a more granular understanding of disparities and facilitating more equitable public health response.



Reproduced from https://www.coronavirus.in.gov/ with permission.
A full list of Indiana's public-facing datasets related to COVID-19 is available here.

- While this problem is not unique to Indiana, MPH and other state agencies have made a commitment to tackling these issues in the coming years.

- **Engage the public.** Not only did MPH engage experts, they also sought feedback from the public on their COVID-19 data sharing and response efforts through widely accessible communication channels, including social media sites. MPH provided updates about changes to data and findings in user-friendly language for maximum transparency. Public-facing dialogue about research, data, and technical infrastructure isn't easy, but it is essential to building trust and forging a path from engagement to co-design with communities. For more on this topic, see the following resources:

  - AISP and the Future of Privacy Forum's "Nothing to Hide: Tools for Talking (and Listening) About Data Privacy for Integrated Data Systems"

  - AISP's "A Toolkit for Centering Racial Equity Throughout Data Integration"


AISP