

# Department of Public Health and Environment Improves Readiness to Combat Health Threats with Integrated Data Ecosystem

## STORY HIGHLIGHTS



### Optimal response to public health crises

Reliable, real-time data fed into an integrated surveillance ecosystem enables fast, precise response and prevention strategies when every minute counts.



### Informed decision-making

Various types of data can be easily accessed, visualized, and analyzed while maintaining security to better understand and respond to public health threats.



### Ever-expanding ROI

Every data source integrated into the data lakehouse increases its value, providing the client with long-term benefits while improving public health.

## ABOUT THE CLIENT

Tasked with disease control and emergency preparedness, this agency develops and maintains statewide data systems that enable faster, more effective public health interventions. It plays a key role in monitoring trends and coordinating responses to protect population health.

## CHALLENGE

In 2019, the Centers for Disease Control (CDC) launched the Data Modernization Initiative (DMI). The goal of DMI is to produce better, faster, actionable insights for decision-making at all levels of public health. The top priority is to strengthen and unify the infrastructure to ensure a response-ready public health ecosystem.

### Disparate systems cause inefficiencies

The state agency's existing infrastructure was comprised of disconnected disease surveillance systems, which required manual processes for data curation, linkage across datasets, and data quality validation. These gaps made it difficult for officials to efficiently and effectively prepare for and respond to public health threats.

### Integrated, response-ready infrastructure enables rapid solutions

They sought to modernize their data architecture to provide public health practitioners with high-quality, real-time, integrated data. That data could then be analyzed to quickly detect and respond to public health threats, understand complex public health problems, target interventions, and improve the health of all state residents.

## SOLUTION

This public health agency partnered with Resultant to achieve the priorities of DMI in a repeatable, scalable manner, creating an ecosystem model that is connected, resilient, adaptable, sustainable, and response-ready.



### Data lakehouse integration

Our first priority was ingesting data from multiple sources into a data lakehouse, a unified platform for storing and analyzing a variety of data types. We applied our experience integrating public health data assets across numerous public and private cloud environments, along with deep in-house expertise in public health and epidemiology, to guide them through this process.



### Key technologies

The platform buildout began with the design and development of foundational architecture, followed by iterating solutions to align with specific needs. The data lakehouse built on the Google Cloud Platform (GCP) utilizes the following tools:

- Google BigQuery to enable large-scale data analytics
- Google Cloud Datastream enables real-time change data capture (CDC) and data replication from multiple sources
- JupyterHub Scheduler to enable users to schedule and run automated jobs



### Efficiency and security

Once the data lakehouse was established and populated, we implemented automated processes to clean and link data in a person-centric way using Resultant's proprietary Probabilistic Record Linkage running on a Databricks compute engine. We deployed a Secure Data Enclave (SDE), or Collaborative Research Environment (CoRE), to enable data sharing and analysis in a secure, isolated environment.



### Custom CDC capabilities

A critical requirement for the agency was the ability to track field-level value changes over time across data sources feeding the data lakehouse, so we leveraged Data Vault 2.0 modeling and real-time data replication as the underlying data modeling methodology. Custom capabilities were developed to support near real-time data replication for data sources that weren't able to natively support real-time replication.



## RESULTS

The agency now has a modern surveillance system data architecture that's integrated, secure, efficient, and scalable. Automation replaces slow, error-prone manual processes, while maintenance of data governance and access policies across departments is streamlined through the data lakehouse governance committee. Most importantly, the agency can now quickly and cost-effectively scale in response to public health crises such as epidemics, pandemics, and other major occurrences.

### Key capabilities and improvements

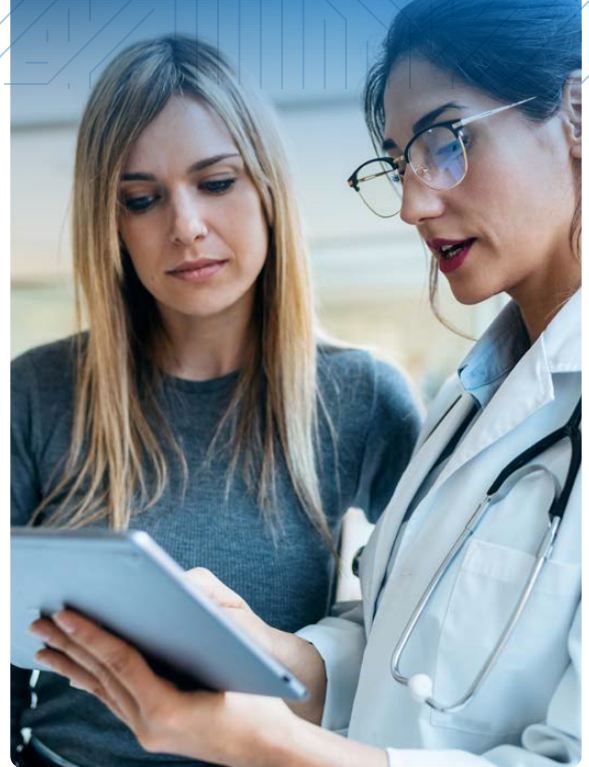
- ✓ Automated near real-time data ingestion
- ✓ Increased data security (row-level permissions)
- ✓ Automated, secure person matching
- ✓ Improved data quality and interoperability
- ✓ Self-service reporting capabilities
- ✓ Data linkage across multiple source systems
- ✓ All data defined in a central data dictionary
- ✓ Standardized data governance
- ✓ Scalable cloud environment

### Secure data access and sharing

Data can be securely analyzed, shared, and used by other systems and divisions within this agency and beyond. Analysts can securely access data, easily publish and locate datasets, utilize their preferred analytic software, and automate the execution of data extraction and analysis via multiple cloud-hosted schedulers.

### What this means for citizens

The modernized data ecosystem is built to scale over time and provides access to a secure, integrated set of critical data sources the agency needs to best prepare for and respond to public health threats. It enables public health resources to more efficiently and effectively access and analyze data that is needed to help improve and protect the lives of all state residents.



## FUTURE

To maximize and accelerate the return on large data asset investments, it's imperative to gain buy-in from data source owners early in the planning process. The benefits mentioned previously will only grow over the life of the health surveillance system as new data sources are added to the data lakehouse, with value determined by the number and breadth of those sources. The more data sources, the better. Resultant continues to provide management and operations (M&O) support for the platform.

**To improve disease prevention and response to public health crises, reach out to Resultant today for a free gap assessment.**