

# Data Lakes Address 21st Century Public Health Challenges

State of Minnesota – Minnesota IT Services

**CATEGORY:**  
Emerging & Innovative Technologies

**CONTACT:**  
Emily Dornfeld  
Director of Communications  
emily.dornfeld@state.mn.us  
(O) 651-201-1011  
(C) 651-485-1354

**INITIATION DATE:**  
March 2020

**END DATE:**  
Ongoing



## Executive Summary

The Minnesota Department of Health (MDH) core disease surveillance and immunization registry systems (MEDSS – Minnesota Electronic Disease Surveillance System, and MIIC – Minnesota Immunization Information Connection) were not designed to handle the data load generated by the COVID-19 pandemic. To address the situation, the agency first moved both systems into the Amazon Web Services (AWS) public cloud in 2020 and then off-loaded non-transactional workloads into a fleet of sophisticated data processing and reporting systems built around data lake technologies.

Over the next 24 months, MDH accelerated efforts in response to COVID-19, driven by the need to integrate disparate data sets and present the results in a consumable format, and to enable data-driven decisions. Outcomes such as the [COVID-19 vaccine dashboard](#), [COVIDaware MN](#) exposure notification system, [vaccination programs](#), [vaccine incentive programs](#), and even case investigation would not have been possible without robust data lakes backing MDH. The systems processed 1.4 million positive COVID-19 cases and 21 million diagnostic tests and tracked over 12 million vaccine doses to-date.

In addition to supporting the development of robust data processing infrastructure, data lakes proved to be a powerful resource for integrating disparate data sets from a myriad of sources and transformed the results into insights that would have been otherwise unavailable.

Using serverless data processing routines, MDH created an environment that drove external reporting ([daily COVID case reporting](#) pages and the COVID-19 vaccine dashboard) and enabled sophisticated data exchange relationships between MDH and external healthcare providers, particularly in immunization efforts. This fueled provider communications and empowered state outreach efforts – over 3.9 million Minnesotans have at least one vaccine dose.

Introduction of the data lake technology made the difference between having high-functioning, integrated systems, or being left with failed critical systems. MDH worked with [Minnesota IT Services](#) (MNIT—the state’s IT agency), vendor partners, and organizations that fed data to MDH, such as health laboratories, healthcare systems, and the federal government – all contributing to Minnesota’s public health pandemic response. MDH uses this technology and architecture as an organizing principle for building out an applied data governance framework and establishing a robust agency-wide data integration and reporting environment. The approach drives strategies that align with the agency’s goal of using data to better inform public health decision-making.

## Idea

### What problem or opportunity does the project address?

The COVID-19 pandemic is the most impactful public health crisis of modern times. Crisis response in the United States has largely fallen onto individual state public health agencies. Each state has approached the situation in their own individual ways. The Minnesota Department of Health (MDH) has led the Minnesota response, supported by Minnesota IT Services (MNIT) as the state IT agency.

The COVID-19 pandemic is unique among modern health crises for its scale and unpredictability. The virus and patterns of human immune response have consistently left the experts guessing how to operate, resulting in rapid changes in requirements and direction, and data products to support public policy decisions. As was the case all over the country, existing epidemiological and immunization technology systems were not designed to handle workloads at this scale. System failure was an existential threat, requiring supplemental technology to be rapidly implemented.

The pace of the pandemic response also required MDH scientists to quickly assimilate new data from multiple sources into meaningful information. Data might come from anywhere at any time, and there were no opportunities to create stable, highly structured repositories such as data warehouses. Data lake technology was well suited to the need, with MNIT MDH cloud data scientists providing the service calls and scripting to quickly integrate incoming data.

### **Why does it matter?**

If a system failed, the MDH Emergency Operations staff conducting the pandemic response would have been flying blind. In effect, they would have guessed the course of the pathogen and its impacts on local populations. We needed robust technology solutions to save lives.

More importantly, the data lakes provided a method for aggregating disparate data in a consumable form that could not be accommodated in core transactional systems. Much of what was aggregated did not exist in traditional database form, including file-based data sources such as Excel spreadsheets. We were able to bring together data and create information products that could not have been effectively created outside of the technology.

### **What makes it different?**

While other states tried to replace their existing technological solutions, Minnesota took a different path by strengthening our core transactional systems and augmenting them with supplemental systems to reduce the load. We focused our core systems on what they are designed for: processing transactions. We then brought in a network of data lakes and serverless application technologies to serve as the backbone of data integration, system integration, and business intelligence functions. Along with massive cloud provisioning of our transactional systems, this approach allowed for exceptional business continuity throughout the pandemic.

### **What makes it universal?**

The approach is transferrable to other state teams, although having a robust cloud services operations (State CIO priority #4) and a capable data management strategy (State CIO priority #9) are foundational to success. It also requires a highly talented team of cloud-centered technical developers.

## Implementation

### What was the roadmap?

The rapid-fire nature of the crisis was antithetical to the concept of a “roadmap.” However, throughout the experience, we well understood that we were paving the way not only for advancement of the MDH IT services portfolio, but we were also accumulating valuable learnings that would fuel State of Minnesota cloud services development for years to come. Development followed an iterative pattern within rapid three-to-seven-day sprints. Successful implementation is apparent through effective system integration, robust reporting features, and well-functioning transactional systems.

### Who was involved?

From inception to maturity, the effort was a close collaboration between MNIT teams and MDH.

- MNIT Teams: IT Cloud Operations, Data Analysis, Application Development, System Integration, and Information Security.
- MDH: Epidemiological, Immunization, and Public Health Laboratory program staff.

This diverse and integrated team of professionals worked tirelessly to advance, adapt, and pivot systems into new forms. Data is sourced through a large, extended community of healthcare delivery and diagnostic testing organizations. The data is aggregated at MDH. Data lakes and the surrounding infrastructure of data processing applications, security systems, and supplemental applications were the difference in making the team successful.

### How did you do it?

MNIT and MDH had the financial resources for this work – everything else was an immense challenge. We had abrupt and aggressive time demands for improvements. The technology was complex and deep, and the staff had little practical experience applying it. We brought in expert vendor resources through AWS, but that only got us part way there. We assembled a core team and proceeded to build expertise within the pandemic crucible. Through time, we leveraged and refined high-functioning architectural patterns that emerged over and over.

*(next page)*

Figure 1: Illustrates the central role data lakes played in disease surveillance efforts.

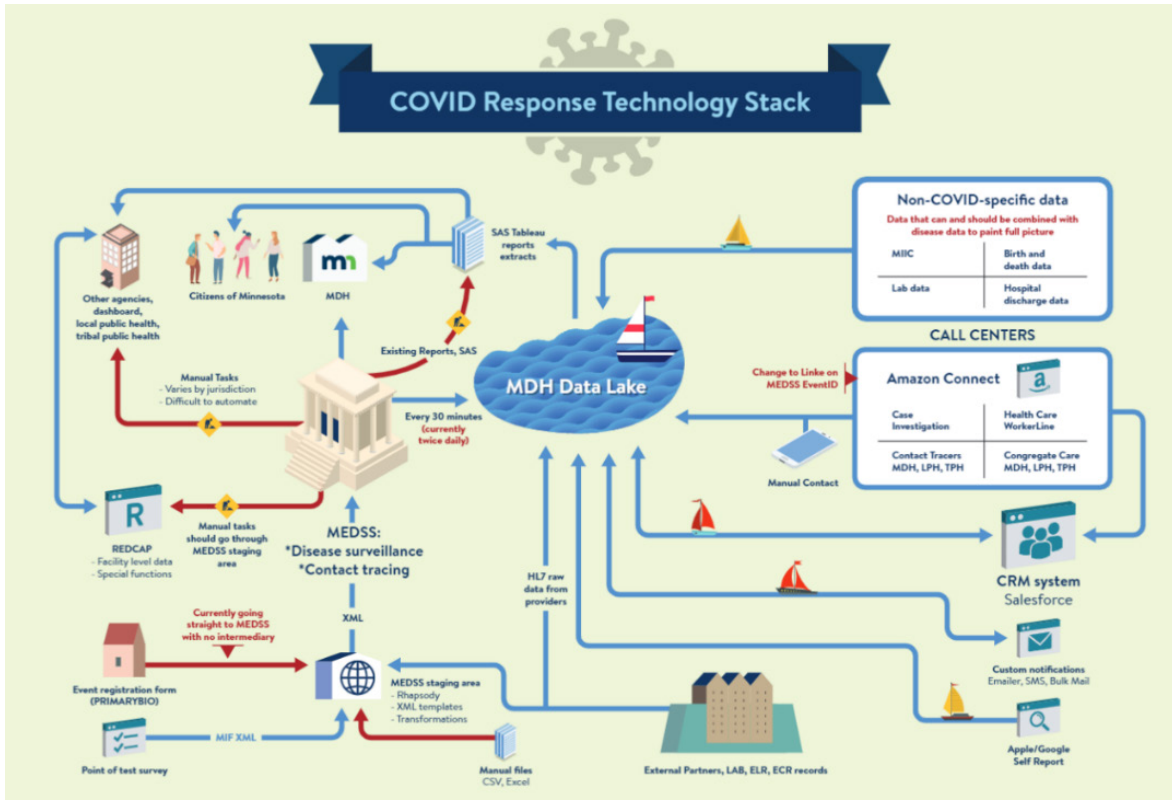
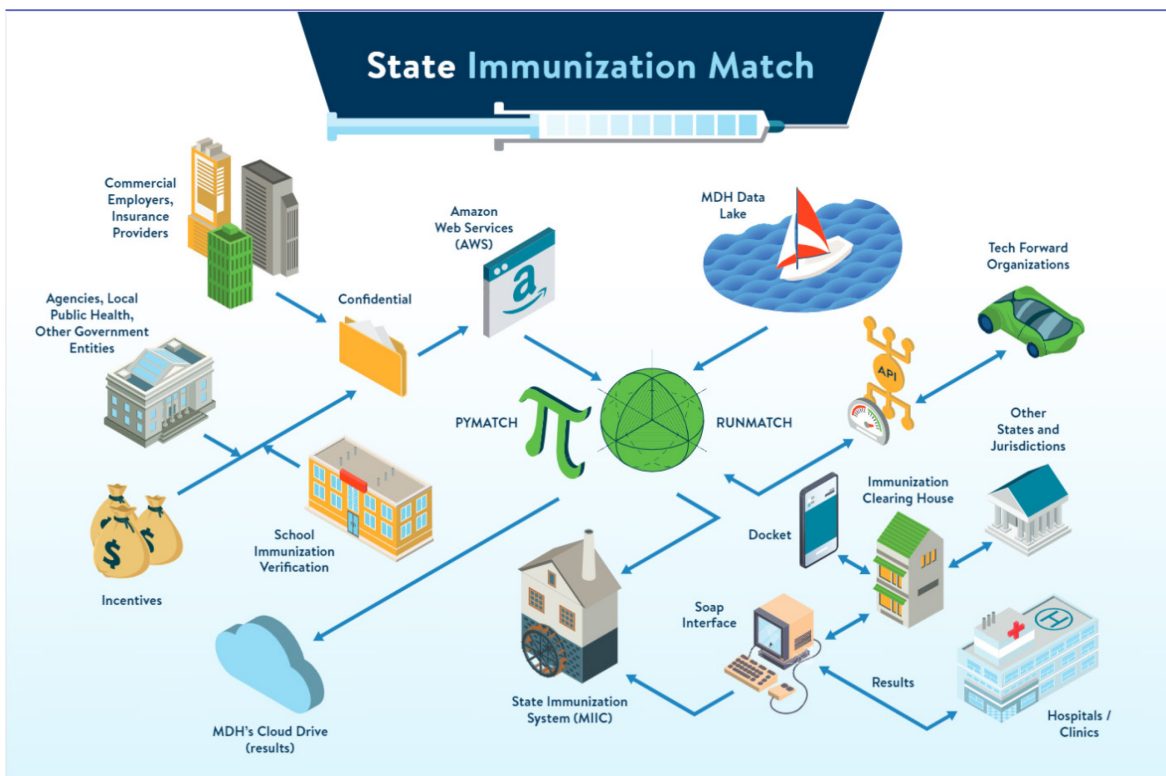


Figure 2: Illustrates how data lakes were used to disperse immunization data into a myriad of information products and technical capabilities.



## Impact

### What did the project make better?

This project enabled the data processing and reporting backbone of Minnesota's COVID-19 response. Once fully implemented, the solution provided a robust, highly available, high-performance, resilient infrastructure that allowed the Minnesota Governor's Office, Department of Health leadership, and Emergency Managers to focus on the policy and operational response to the crisis. It also proved to be flexible. We added new capabilities to meet new demands as the pandemic changed.

Additionally, our network of data lakes supported HL7-based confirmed laboratory case intake and processing (an international standard for healthcare data) and also served as a primary point of integration with CRM-based call center operations. When the Governor's Office requested a vaccine incentives program, the data processing infrastructure was already in place to support this new family of applications.

Our epidemiological and immunization systems were stable but ill-equipped to handle the massive workloads thrown at them. They also were poorly positioned to support the additional diverse capabilities that the pandemic response demanded. By off-loading major data exchange, integration, and reporting workloads from the core transactional (and legacy reporting) systems, we armored them with focused upgrades and created stability that would otherwise be elusive.

While these systems exist primarily behind the scenes, their outcomes reached millions of Minnesotans and others across the nation seeking information on the status of the COVID-19 emergency in Minnesota.

### How do you know?

Quantitative data on the effort is difficult to come by since in the midst of the pandemic, it was difficult to measure the impact of our work. The team was simply trying to get through the experience on a day-by-day basis, although it was clear to the technologists and business partners who worked with these systems previously that they would have ceased to function had the supplemental systems not been put into place.

MDH staff were at first skeptical of the approach, but ultimately came to embrace data lake technology whole-heartedly ([see this video](#)) and have since come to recognize both the impact and potential of the technology well beyond the COVID-19 pandemic response.

### What now?

The project was transformative for public health operations in Minnesota. Our data lake experience is transforming new application areas and is being institutionalized as our primary method for data integration, aggregation, and reporting. The framework provides us with a real-world basis for developing actionable data governance processes and creating data-driven solutions. A crisis has become our opportunity to move a modernization agenda in ways that would have been difficult to achieve otherwise.