Colorado data modernization: Assessment to action

CONTRIBUTOR: *Paul Gillenwater, Communicable Disease Technology Manager, Colorado Department of Public Health and Environment*

CSTE
Stories FROM THE Field

CATEGORY: Epidemiology and Laboratory Capacity (ELC)	CATEGORY: Enterprise Approach to Data Systems Modernization

To meet the public health response needs, including case reporting and contact tracing, the Colorado Department of Public Health and Environment (CDPHE) had to invest significant time and resources to enhance old systems, integrate data between systems, and develop new solutions for case management and data sharing.

The public health system in Colorado is decentralized, made up of the state health department and 56 local health departments, but disease surveillance systems are shared between state and local entities. The COVID-19 pandemic stretched Colorado's surveillance systems to their limits. To meet the public health response needs, including case reporting and contact tracing, the Colorado Department of Public Health and Environment (CDPHE) had to invest significant time and resources to enhance old systems, integrate data between systems, and develop new solutions for case management and data sharing.

The "What"

CDPHE recognized the need for a comprehensive roadmap to modernize and consolidate their disease surveillance systems. They engaged with a consulting firm in late 2021 to assess existing surveillance infrastructure and make recommendations for modernization. Two projects in line with CDC DMI goals were recommended. The first was to consolidate CDPHE's disease surveillance systems into a more comprehensive and flexible platform. The second was to create a Data Lakehouse to house, connect and share all available data relating to disease control.

Colorado's current disease surveillance system landscape is spread across 11 distinct systems with many additional helper applications and processes that facilitate their operation. Many of these systems cover a specific disease group (e.g., tuberculosis, lead, zoonotics, STI/HIV), which effectively isolates data for these conditions and often requires integration with Electronic Laboratory Reporting (ELR) and the maintenance of technology that is more than 20 years old. Their core surveillance systems, the Colorado Electronic Disease Reporting System (CEDRS) and ELR, were homegrown. While the systems have been updated over time, they retain their basic original design and functionality.

CDPHE identified a few key areas to focus on when considering a replacement. Can a replacement platform account for the surveillance needs of all reportable conditions? Is it flexible and configurable? Does it facilitate automation? Is it scalable and able to handle major public health responses? After surveying all states

and their current disease surveillance systems and considering a new build from scratch, EpiTrax rose to the top. CDPHE communicated with the EpiTrax consortium, received access to source code and built out their own EpiTrax "sandbox" environment for further evaluation. After additionally interviewing the jurisdictions that currently use EpiTrax, they ultimately chose EpiTrax as their future surveillance platform. EpiTrax is highly configurable with the form builder component, relies on a high level of automation with the rules engine in the Electronic Message Staging Area, can cover all reportable conditions in Colorado, and is scalable and secure with cloud hosting.

CDPHE wrote a request for proposals for a vendor to assist Colorado with the transition and finalized a contract with their chosen vendor in November 2022. The following month, they began a 20-month contract to transition the patchwork of homegrown disease surveillance systems to EpiTrax and began to plan for an initial launch of core systems (which, at minimum, includes CEDRS, ELR, contact tracing and outbreak management) in December 2023, with a secondary launch including all remaining non-core systems by July 2024. They launched their first end-to-end instance of EpiTrax in early March 2023 and made it accessible to members of state and local health department workgroups focused on helping with the transition.

The "So What"

This transition will have a significant impact on Colorado's ability to conduct disease surveillance and provide public health services. Already, the transition has entailed conversations about how to plan for and make the most out of a new platform that will increase the automation of ELR and allow for the implementation of Electronic Case Reporting (ECR), leading to increased communication and cooperation with local health jurisdictions, hospitals, Health

Information Exchange systems and internal staff. CDPHE has entire teams of people who have spent their careers becoming experts in manually guiding ELR messages into CEDRS. The transition will be an opportunity for those teams to apply their expertise in a new way by creating rules for the Electronic Message Staging Area to rely on for automation. Local public health jurisdictions are getting excited about new possibilities for accessing surveillance data, and CDPHE has committed to creating an API for their use, which will allow them to automate their processing and use of surveillance data for the first time in Colorado.

The transition also necessitates change, which will be difficult for some users, infection preventionists (IPs) in particular, as they adapt to a new system that doesn't allow for facility-level access. In the long run, this change will allow Colorado to focus their efforts toward ensuring that data is complete and accurate at the source (ELR and ECR), instead of relying on IPs to fill in the gaps when data is incomplete. This will lead to closer relationships with technical staff responsible for sending data electronically and will ease the reporting burden on IPs that Colorado has relied upon for so long.

The "Now What"

Colorado is eager to become a contributing member of the EpiTrax consortium. They have a list of enhancements they would like to devote resources towards once they have fully transitioned, including building out text messaging functionality, data quality and automation tools, and adding privileges that will allow our partner agencies to have increased access to their own data. This project has resulted in new expenses that they have

not historically had to account for in Colorado, as well as vendor-managed hosting, maintenance and development of new functionality.

While working on the transition to EpiTrax and after reviewing internal resources and concluding that Colorado's current infrastructure could not be used, CDPHE wrote a Request for Proposals for a Data Lakehouse. The RFP resulted in selecting a vendor

that has years of experience with building data vaults, warehouses, Data Lakehouses, and working in public health departments. The consulting company Colorado engaged set a foundation for the state to move toward and was the foundation of the RFP the vendor responded to, which included identity resolution services and a master person index that will live in the Data Lakehouse. The vendor has completed similar work in Indiana, during which they built an application for internal and external users to access information through an application known as CoRE (Collaborative Research Environment). Colorado will also use a CoRE environment.

Throughout the updating process, data governance has been a major focal point to ensure the correct data is accessed by the correct entities. This level of data governance in the Data Lakehouse will be achieved with the use of identity resolution services, a master person index and CoRE. Colorado's implementation started in January 2023 and will be completed by the end of June 2023. The final implementation is slated to include EpiTrax, immunization, newcomer/refugee, HIV/STI and vital statistics data. The implementation plan will be completed in four different phases. The first phase, which will be completed by the end of August 2023, will focus on bringing immunization and newcomer health data into the Data Lakehouse.

With the implementation of a Data Lakehouse, Colorado will have the ability to address public health concerns and respond to future incidents. During the COVID-19 pandemic, the siloed nature of Colorado's application data made it time-consuming and difficult to answer urgent and high-consequence questions. For example, the pandemic response necessitated analyses of COVID-19 vaccine effectiveness, but they found that the immunization database and disease surveillance database did not link effectively.

The Data Lakehouse will make solving such issues trivial in the future. CoRE will greatly improve data access for internal and external stakeholders, including local public health agencies. Instead of extracting data from applications manually using pre-built reports, stakeholders will have more complete and automated access to the data they need.

The Data Lakehouse will also greatly reduce time spent cleaning, completing and integrating data, allowing for more meaningful work such as identifying health equity concerns or modeling with predictive machine learning algorithms. Colorado's future disease surveillance data infrastructure could allow public health to get ahead of an outbreak, rather than simply collecting data from those that have already happened. The Data Lakehouse also provides the building blocks the state will need to meet future infrastructure needs. As a cloud-based solution, it is scalable, it allows for improved data security and access management, and it provides the opportunity to train analytic staff with modern technology resources. The Data Lakehouse will give Colorado the ability and flexibility to make data-driven decisions moving forward.

Since the onboarding of the Data Lakehouse vendor in January 2023, CDPHE has been finalizing the implementation and architecture plans. Overall, they are taking a phased approach, which involves incorporating data from one or two sources at a time. In each case, they work closely with corresponding data teams to make sure the new data model will suit their needs effectively. This process has resulted in connecting teams that would have otherwise remained siloed. CDPHE hopes the Data Lakehouse will be the beginning of a cultural shift in how the department views its data: not as the privileged domain of a few experts, but as a resource to be used as effectively as possible.