Swimming in the data: Utilizing a data lake for eCR data quality in Minnesota

CSTE

Stories FROM THE Field

CATEGORY: **Epidemiology and Laboratory Capacity (ELC)**   CATEGORY: **Electronic Case Reporting**

Despite electronic case reporting (eCR) speeding up the ability for staff to receive reports, Minnesota Department of Health noted issues with data quality and dropped feeds. Leveraging their data lake, MDH identified critical data elements in eCRs and extracted them to conduct data analyses. Findings were communicated to submitters to improve data quality and MDH was able to more quickly identify and repair dropped feeds.

The "What"

When electronic case reporting (eCR) became viable in 2020 for COVID-19 reporting, the Minnesota Department of Health (MDH) quickly implemented techniques to process the eCRs into their surveillance system and attach a readable format. This sped up the ability for program staff to receive reports and extract relevant information, but it didn't help with monitoring data quality, dropped feeds, or quickly identifying issues within certain submitters due to the inability to review more than one raw message at a time, a process necessary for onboarding of new submitters to ensure data quality even in the midst of a pandemic.

Data was already being placed into an archiving data lake, but not utilized. Working on an interdisciplinary team, MDH identified important data elements within an Electronic Initial Case Report (eICR) and Reportability Response (RR) file to extract and place into formatted tables. They saved valuable processing resources by creating the data quality reporting outside of the surveillance system, which allowed them to investigate or create adhoc reports more easily.

The "So What"

Improved eCR ethnicity from one submitter rose from 2% to 85%: MDH identified a specific healthcare organization (HCO) with extremely low ethnicity completeness. They noted a huge discrepancy compared to their electronic laboratory data and reached out to the HCO, which was able to identify missed mappings.

Able to identify dropped feeds at least four times in under two days and before central platform alert went off: In the nine months, MDH identified four instances of dropped feeds between the HCO and APHL, delaying case reporting not only to MDH, but other public health agencies too. With early detection of dropped feeds, MDH would alert external partners, which sped up fixing the issue and decreasing a mass flux of backlog reports flooding their surveillance system. They have established protocols to alert surrounding states that would be impacted by an HCOs dropped feeds to ensure information is disseminated.

→ More stories are available at **stories.cste.org**

Minnesota has been able to push the conversation at the national level and get buy-in from all jurisdictions that the national onboarding team can confidently promote removing triggering from resolved problems.

Contributed to national conversations related to eCR issues: MN has been able to communicate issues related to old active problems or diagnosis-triggering eCRs that do not need to be reported to PHAs. They've been able to push the conversation at the national level and get buy-in from all jurisdictions that the national onboarding team can confidently promote removing triggering from resolved problems. MDH was able to be a part of the pilot for timeboxing problems and diagnosis that are left on patient charts with the data lake to demonstrate success in timeboxing and promote that 30-day window used in timeboxing is too broad. The ability to alert on dropped feeds earlier helped in establishing a national workgroup for the need/creation of a PHA dashboard on the APHL platform for use in monitoring dropped feeds, creating reports related to volume, ELC grant metrics, and onboarding statuses.

The "Now What"

The data lake set up has proven successful and a solution for other reportable conditions to access their data or send their eCRs to a different surveillance system. This is fundamental in going from infectious disease case reporting to any reportable condition reporting.

Health equity lens: Many healthcare organizations are now collecting health equity questions in the patient's social history that is coming through in eCRs. As health equity is a priority for MDH, review and use of this available data could help in prioritizing state resources or better ways to target public health communications.

Promote better integration with NIOSH at the healthcare organization first rather than just at

the public health agency: MDH knows some EMR vendors store occupational data as text rather than coded. When this happens, it doesn't meet the transmission standards for eCR and occupational data will continue to stay inoperable to public health. We need conversations at the national level for NIOSH coding to be available not only for public health, but healthcare organizations too.

Continue data lake use for communicating with onboarding and production partners to resolve issues earlier and build out these processes to be more automated: Building out monitoring volume of eCRs received against lab reports improves the ability to track data quality issues for both ELR and eCR. This will contribute and promote solutions at the national level by quantifying issues with the data lake.