

CDPH COVID-19 After Action Report Chapter 19 – Data and Reporting

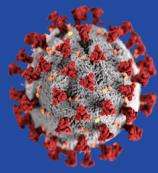
Chapter 19 Data and Reporting



Version History

Version #	Date	Notes
0.1	6/30/2023	First Draft submitted to CPR team
0.2	8/3/2023	Final Draft revised per review by CPR Team
0.3	2/7/2024	Final Draft revised per Expert Review and CPR Leadership review
1.0	7/1/2024	Final revised per CDPH Directorate review
1.1	1/9/2025	Final rebranded

DRAFT

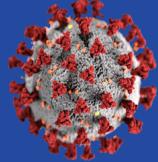


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Table of Contents

19. Data and Reporting	19-1
Chapter Summary	19-1
Overview	19-1
Main Strengths and Successes.....	19-9
Main Challenges and Lessons Learned	19-14
Analysis of Activities	19-22
Data Teams and Workstreams.....	19-22
Data Reporting	19-24
Data Submission	19-34
Data Quality and Remediation	19-38
Informatics and Data Integration.....	19-41
Data Dashboards and Visualization.....	19-45
Data Workforce	19-50
Equity	19-53
Data & Technology	19-54
Communications.....	19-58
Internal	19-58
External	19-60
Workplan	19-63



19. Data and Reporting

Related Public Health Emergency Preparedness and Response Capabilities:

Public Health Surveillance and Epidemiological Investigation; Information Sharing.

Related CDPH AAR chapters: Epidemiology and Surveillance; Enterprise Technology; Testing; Vaccines; Medical Surge.

In this chapter, some abbreviations may be used interchangeably with their respective full spellings for ease of reading.

Chapter Summary

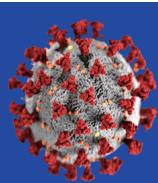
Overview

This section provides a high-level overview of milestones and activities related to this chapter.

Across the country, reportable disease conditions are defined by the CDC and expanded by state laws. Laboratories (labs) and providers must report these disease conditions to facilitate public health surveillance, which is the ongoing collection, analysis, and dissemination of public health-related data. Public health surveillance requires close collaboration between the data managers and informaticians who collect the data, and the epidemiologists who use model, analyze, and investigate the data to determine outbreaks, trends, public health actions to control and prevent disease, and other activities.

As the COVID-19 pandemic accelerated in late spring 2020, CDPH was confronted by increasing needs for data. Case reporting and laboratory test results were the first data streams that quickly increased to high volume beyond normal reportable diseases during the COVID-19 pandemic response. Case data was derived from local public health investigation and lab test results after they are processed. Test and case data were some of the important metrics used to track, monitor, and respond to the COVID-19 pandemic. Additional data streams, such as hospitalization (including ICU admission), hospital census and capacity, death, vaccine and therapeutics receipt, inventory, distribution, and utilization, outbreaks, and other supplemental data were also added throughout the pandemic response.

CDPH's data and reporting work evolved significantly over the course of the response and involved many teams, who worked with different datasets at



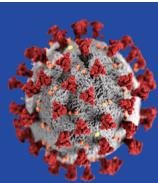
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

different stages throughout the data lifecycle. These stages include data receipt, cleaning and remediation, processing, dissemination, visualizing, and reporting. **Figure 1** depicts major data streams that CDPH reported on during the pandemic response.

Figure 1: Major Data Streams and Associated Systems

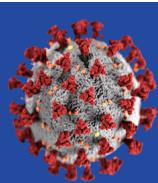
Data Type	Associated Systems	Description	Covered in this Chapter?
Test	CalREDIE, COVID-19 Reporting System, Snowflake	Test data includes all test results, including positive, negative, and probable, reported by laboratories. Test data is used to monitor testing volumes, testing equity, and other testing trends.	Y
Case	Case registry dataset, CalREDIE, CCRS, Snowflake	Case data is processed test result data that has been recoded based on lab results and other criteria. Case data focuses on positive and probable test results and is used for contact tracing, case investigation, and analyzing outbreaks.	Y
Death	CalREDIE, Snowflake	Death data is based on death investigations and includes demographic, clinical, and epidemiological information collected by local health jurisdictions (LHJs) to confirm	Y



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Data Type	Associated Systems	Description	Covered in this Chapter?
		whether that a death is COVID-19 associated.	
Vaccination	CAIR2	Vaccination data includes all COVID-19 immunizations submitted by providers. Vaccination data is used to monitor vaccine uptake, vaccine equity, and other vaccination trends. COVID-19 vaccination data is discussed in this chapter in the context of integration with other data streams. For data related to vaccine allocations, inventory, and distribution, see the Vaccines chapter in this AAR.	Y
Hospitalization	N/A	Hospitalization data includes data on bed capacity, intensive care unit (ICU) capacity, and staffing capacity. Hospitalization data is used to prepare for medical surges and monitoring other trends.	Y
Skilled Nursing Facility (SNF)	Survey 123	This data stream is managed by CDPH's Center for Health Care Quality (CHCQ). See the	N



CDPH COVID-19 After Action Report

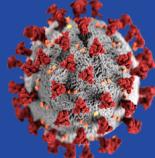
Chapter 19 – Data and Reporting

Data Type	Associated Systems	Description	Covered in this Chapter?
		Infection Prevention chapter in this AAR.	
Resource Requesting	Public Health Ordering System (PHOS)	This data stream is managed by CDPH's Medical and Health Coordination Center (MHCC). See the Resource Requesting and Public Health Ordering System chapter in this AAR.	N
Therapeutics	Health Partner Ordering Portal (HPOP) and Tiberius (federal systems)	This data stream is handled by the Therapeutics Task Force. See the Therapeutics chapter in this AAR.	N

- For a discussion of major systems from an Enterprise Technology perspective, see the Enterprise Technology chapter in this AAR.
- For a discussion of Data Modeling, see the Epidemiology and Surveillance chapter in this AAR.
- For a discussion of CDPH's additional COVID-19 surveillance activities, including sero-surveillance, wastewater surveillance, and others, see the Epidemiology and Surveillance chapter in this AAR.
- CalCONNECT is discussed in this chapter only in the context of its relationship with CalREDIE and death data. For a discussion of CalCONNECT, see the Contact Tracing chapter in this AAR.

Data Reporting Initiation: Spring 2020 – Fall 2020

Historically, CDPH's public disease reporting occurs on an annual basis and garners little publicity. In general, during outbreaks epidemiologists at the state level would collect data, wait until the outbreak was over, and then publish the data. However, the size and scope of the COVID-19 pandemic created new expectations to collect and report on data. Californians wanted access to current, accurate, and relevant data around the clock; the California media



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

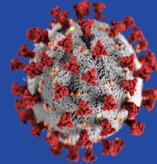
wanted to report on data every day; and California's policymakers and leaders wanted real-time data to make and adjust policy decisions and interventions.

In January and February 2020, the federal government (CDC) approved and conducted testing and CDPH counted cases using whiteboards and Excel. This quickly changed over the course of several weeks. Testing capability began to shift to the state and local public health labs, testing volumes began to increase, and more providers and labs later began offering tests. COVID-19, the new disease condition, was added to CalREDIE.

CDPH began reporting COVID-19 data seven days per week to the CDC, State leadership, and the public. These new reporting expectations and requirements were unprecedented, the volume was exponentially higher than any prior reportable diseases, and CDPH's data and reporting teams struggled to adjust. CDPH was under intense scrutiny and pressure to meet these data and reporting expectations on very short timelines.

These high expectations were in direct conflict with the realities of CDPH's legacy data systems. CalREDIE, the State's disease surveillance system, had long been underfunded and was inflexible, slow, difficult to configure, and lacking in functionality. As testing volumes increased dramatically, CalREDIE became increasingly unstable, and CDPH began working on both temporary and permanent infrastructure and system improvements. However, these improvements took time. In the meantime, CDPH established ad hoc, manual reporting processes using Excel and PowerPoint, and separate processes to receive data from local health departments not using CalRedie such as Los Angeles and San Diego, which comprise almost 1/3 of California's population.

As explained further in this chapter, data and reporting SMEs described this first phase as being defined by "chaos," "crisis mode," and "reactivity." CDPH was inundated with data requests and was initially unable to proactively improve data reporting quality, timing, and thresholds. As a result, teams devoted considerable time to researching and investigating data discrepancies that would have no policy impact. Additionally, data quality concerns resulted in month-long delays in releasing data by race, ethnicity, and gender. Eventually, early fall 2020, CDPH began reporting case and death data by these equity metrics. With the implementation of the COVID-19 Case Reporting System (CCRS) gateway for receipt of electronic lab data and the Snowflake data warehouse in Fall 2020, CDPH stabilized its data reporting infrastructure and systems and began to automate data reporting.

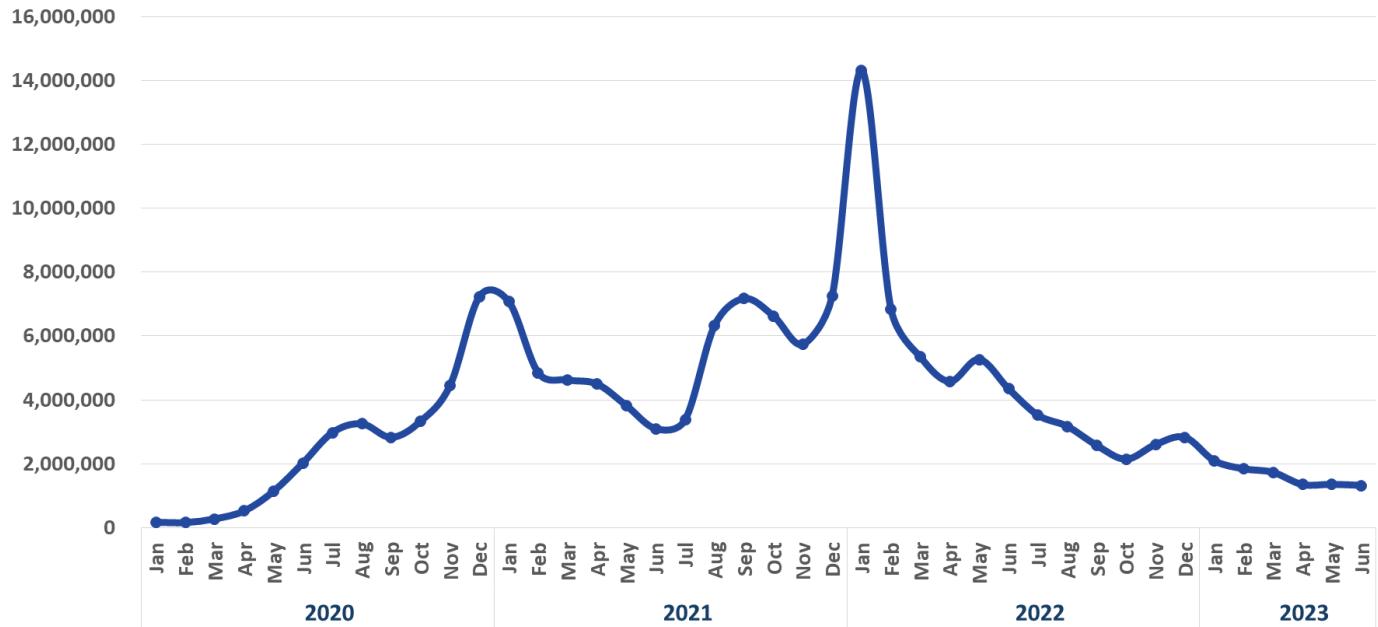


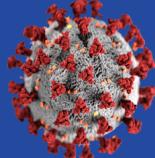
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Figure 2 depicts the total count of electronic lab results received and processed by CalREDIE, CDPH's statewide electronic disease surveillance system, from January 2020 through June 2023.

Figure 2: Total Count of Electronic Lab Results Processed by CalREDIE (January 2020 – June 2023)



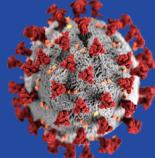


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Timeline and Key Milestones

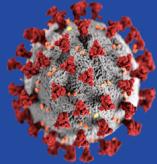
2020	
Winter 2019/2020	<ul style="list-style-type: none">• Mid-January: New COVID-19 disease condition added to CalREDIE• January – February: Early data and epidemiology teams formed within the Immunization Branch
Spring 2020	<ul style="list-style-type: none">• Early March: CDPH started reporting line list COVID-19 case data to CDC• March 9: California Code of Regulations Title 17 updated to add COVID-19 to list of reportable conditions• March 18: CDPH requested labs to start reporting negative (in addition to positive) COVID-19 test results• March: State COVID-19 dashboard (also known as "Governor's Dashboard") went live• April: "Lane 6" data reporting team created• March – April: Decision made to use Cal OES as source of report for COVID-19 death data
Summer 2020	<ul style="list-style-type: none">• May: Lab Testing Metrics Application went live to estimate case counts• May: County Data Monitoring webpage created• May – June: Commercial testing became more widely available• June: CDT published unsuppressed COVID-19 data on State open data portal• Late July: As a result of data system failures, no COVID-19 test results are reported for several days, leading to inaccurate case rates• Late July: CDPH expanded data reporting requirements for labs and providers to include race and ethnicity• August: First U.S. case of COVID-19 reinfection reported• August: Blueprint for a Safer Economy (plan for reopening based on case metrics) webpage created• August: Federal guidance expanded data reporting requirements for labs and providers to include race and ethnicity• August: CDPH assumed responsibility for publishing suppressing COVID-19 data on CalHHS open data portal
Fall 2020	<ul style="list-style-type: none">• September: All COVID-19 data dashboards suppressed in accordance with de-identification guidelines• October: New COVID-19 Case Reporting System lab results gateway and Snowflake data warehouse went live• November: New Data Processing and Informatics Section (DPI) created and Lane 6 discontinued



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

2021	
Winter 2020/2021	<ul style="list-style-type: none">• December: Lab Testing Metrics Application discontinued• Dec. 13: CalREDIE auto-processing turned on for positive lab test results• Late December: Vaccines arrived in California, inaugurating a new vaccine inventory and vaccination data stream
Spring 2021	<ul style="list-style-type: none">• March: Source of report for COVID-19 death data changed to CDPH• March: CDPH began reporting cases by episode (onset) date in addition to report date
Summer 2021	<ul style="list-style-type: none">• July: Hospitals required to report line list discharge data per AFL 21-25• July 7: California Code of Regulations Title 17 updated to require labs to report COVID-19 whole genome sequencing data• August: CDPH implemented a one-day lag in data quality and clean-up processes
Fall 2021	<ul style="list-style-type: none">• September: CDPH began reporting case rates by vaccination status, reflecting the integration of these two datasets
2022	
Winter 2021/2022	<ul style="list-style-type: none">• December – February: Omicron Surge
Spring 2022	<ul style="list-style-type: none">• Spring: CalREDIE configured to automatically detect reinfections
Summer 2022	<ul style="list-style-type: none">• June: CDPH assumed management of State COVID-19 dashboards and associated data pipelines
2023	
Winter 2022/2023	<ul style="list-style-type: none">• February 28: California's State of Emergency for COVID-19 ended
Summer 2023	<ul style="list-style-type: none">• June: MHCC deactivated from the COVID-19 pandemic response



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Main Strengths and Successes

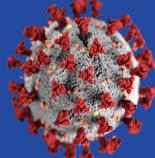
This section describes the Main Strengths and Successes, including findings and corrective actions, related to this chapter. Further elaboration and a more detailed discussion of these strengths and successes can be found in the Analysis of Activities section.

1. CDPH created new data systems, improved legacy data systems, and automated data reporting processes to receive, clean, process, disseminate, and report on unprecedented volumes of COVID-19 data.

CDPH made significant progress to establish robust data and reporting solutions throughout the pandemic response to process large data volumes, integrate various data streams, meet new requirements, and provide accurate, timely, and high-quality data. For other reportable disease conditions, CDPH was accustomed to a weekly reporting cadence to the CDC and annual website updates. However, in early 2020 CDPH teams found themselves struggling to meet changing reporting requirements, evolving data streams, and unprecedented data volumes. Initially, legacy system limitations and a lack of data centralization led to data quality problems and incomplete data reporting that required significant manual work. The implementation of the COVID-19 Case Reporting System gateway for lab data and a Snowflake data warehouse in fall 2020 represented a major milestone in automating COVID-19 data reporting. CDPH achieved another milestone in 2022 when it assumed ownership of the data warehouse. One leader noted, “it’s amazing how much innovation we did in a short amount of time in terms of data management, pipelines, pathways, and exchange.” SMEs agreed that CDPH should continue working on improving its data standardization, automation, and innovation across all its programs.

Finding/Corrective Action: CDPH should continue to pursue its data and reporting systems and process improvements in pursuit of improved data standardization, automation, integration, and sharing. (ID: Data and Reporting 1)

2. CDPH established a new Data Processing and Informatics section to better standardize, integrate, and centralize COVID-19 data streams.



CDPH COVID-19 After Action Report

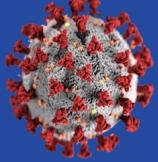
Chapter 19 – Data and Reporting

With the establishment of the new Data Processing and Informatics section in Fall 2020, CDPH achieved a significant milestone in data standardization, centralization, and integration. The newly formed team maintained the “source of truth” for tests, cases, and other COVID-19 data, and also created integrations across disparate datasets for other teams to use in their analysis. This team of “data stewards” provides automated, curated integrations involving test case, case data, vaccination data, death data, and hospitalization data. For instance, case data and vaccination data are stored in separate systems, but the data informatics team integrated these datasets so that epidemiologists could research topics such as breakthrough infections in individuals already vaccinated against COVID-19. According to SMEs, in anticipation of an emergency response it is never too early to start planning data centralization and how to translate raw data into useable formats.

Finding/Corrective Action: CDPH should maintain its centralized team dedicated to overseeing all data streams, reporting, and integration.
(ID: Data and Reporting 2)

3. CDPH successfully corrected early decisions about data and reporting sources, leading to improved data quality and accuracy.

In early 2020, the State made two rapid, significant decisions for data reporting that created unintended consequences. At this point, the California Department of Technology (CDT) was charged with leading the State’s COVID-19 data reporting, and instead of using CDPH’s CalREDIE system as the source of death data, it was decided that Cal OES would provide death data gathered from a variety of different sources other than public health reports. Secondly, instead of reporting COVID-19 cases by episode date (a standard practice in public health surveillance), California decided to report cases by report date. The nuances of these decisions were difficult to understand and not immediately clear, but negatively impacted the data quality that would be reported for over a year. CDPH’s technical experts, including data and reporting team members and epidemiologists, soon realized that these reporting decisions were leading to lower-quality data and began advocating for change at multiple levels within the response. While it took until 2021, this effort was successful—CDPH took over reporting death data and began reporting cases by episode date in addition to



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

report date. According to one data and reporting SME, “those issues were the biggest sources of frustration, and both of them got fixed, which was a success.”

Finding/Corrective Action: In the future, CDPH should be the source of report for all public health data. (*ID: Data and Reporting 3*)

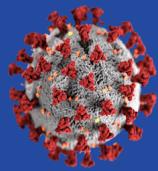
Finding/Corrective Action: Reportable disease conditions should be reported by episode date. (*ID: Data and Reporting 4*)

Finding/Corrective Action: Early in any response, CDPH should approach data and reporting decisions strategically with the awareness that once established, metrics and data sources can be very difficult to change. (*ID: Data and Reporting 5*)

Finding/Corrective Action: CDPH should work with Cal OES to improve death reporting during future emergency responses. (*ID: Data and Reporting 6*)

4. The pandemic response resulted in improved data literacy, a recognition of the criticality of public health data systems, and appreciation for the complexity of data engineering.

The use of data in California’s COVID-19 pandemic response has led to a greater recognition among stakeholders of the importance of data and its role in informing public health policies to mitigate disease and address inequities. The increased visibility of CDPH’s data systems has raised awareness of the need to continue investing in public health systems and fostered greater data literacy amongst stakeholders. It has also created a deeper understanding of the complexity involved in behind-the-scenes data engineering and management. Modernization projects for several key data systems, including CalREDIE and CAIR2, are currently underway, and CDPH is soliciting feedback from a wide variety of stakeholders. “There’s a new recognition of how key the data is to everything, and the process of turning raw data into actionable policy,” one SME noted. Many agreed that it will be crucial to sustain this momentum and continue to fund and improve public health data systems for more timely public health awareness and interventions for all reportable diseases, and in preparation for the next emergency response.



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Finding/Corrective Action: CDPH should continue to reinforce the importance of maintaining and funding public health data systems to improve response. (*ID: Data and Reporting 7*)

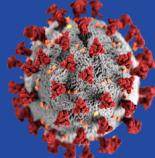
5. **Throughout the pandemic response, communication and coordination between various data and reporting teams improved dramatically.**

As the response progressed there was a significant improvement in communication and coordination between various data and reporting teams. These teams encompassed different groups that worked with diverse data streams at various stages of the data lifecycle, ranging from initial receipt to processing, dissemination, visualization, and analysis. Strong partnerships were forged between data, reporting, informatics, epidemiological, and the CalREDIE program teams, in conjunction with CDPH's Information Technology Services Division (ITSD). Although each team had distinct roles, they all shared a mutual dedication to maintaining data integrity and implementing solutions to enhance California's data reporting capabilities. Additionally, communications around data governance also improved significantly. Now, when changes are made in one major system (such as CalCONNECT), other system teams are notified to avoid downstream complications. Overall, data and reporting teams emphasized the intense camaraderie, ingenuity, and dedication of their response colleagues. These tight partnerships and improved communications helped break down silos and bridge gaps across disciplines.

Finding/Corrective Action: CDPH can sustain these improved communications and partnerships through regular meetings and ongoing coordination, and should consider adding a Data Branch in future data-intensive ICS structures. (*ID: Data and Reporting 8*)

6. **CDPH and its partners shared COVID-19 data with the public through innovative, visually appealing dashboards and the Open Data Portal, empowering Californians to access and analyze information on their own.**

Prior to the pandemic, most public disease reports were static tables or graphics, often with data from months or years prior. The State established multiple, comprehensive dashboards to visualize data for public consumption. Starting in March 2020, California began to report case, death, and hospitalization data on public data dashboards. These



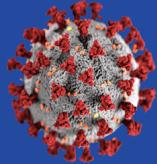
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

dashboards presented data, charts, and tables that became more visually appealing and user friendly over time. In addition to these dashboards, California also made raw COVID-19 datasets available via its Open Data Portal. This portal enabled anyone to download raw data sets for their own analysis and research. Over the course of the pandemic CDPH, CalHHS, and State leadership released many iterations of the dashboards to add new datasets, integrations, and improve data privacy, accuracy, and integrity. The COVID-19 dashboards and the Open Data Portal are seen as unequivocal successes that empowered Californians to obtain up-to-date information on the latest COVID-19 facts and trends.

Finding/Corrective Action: Public data sharing via dashboards and the Open Data Portal represent a successful model, and CDPH should invest in the staff and systems to leverage this model for other significant reportable diseases and future responses. (*ID: Data and Reporting 9*)

DRAFT



Main Challenges and Lessons Learned

This section describes the Main Challenges and Lessons Learned, including findings and corrective actions, related to this chapter. Further elaboration and a more detailed discussion of these challenges and lessons learned can be found in the Analysis of Activities section.

7. CDPH's legacy data systems were not equipped to handle COVID-19 data reporting demands.

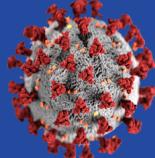
The pandemic highlighted the criticality of having modern, robust technology and data systems needed to deploy public health actions and make rapid public health policy decisions. In early 2020, as requests for more and different types of data increased, and the sheer volume of incoming data increased, it quickly became clear that CDPH's legacy data systems were not designed or equipped to handle these requests. Consequently, CDPH established ad hoc data and reporting channels that were later very hard to change even if they were imprecise. The systems in place, especially CalREDIE, were not designed for a pandemic of this magnitude. Over the course of the pandemic, CDPH significantly improved its technology infrastructure, yet despite these improvements, its public health systems remain vulnerable and fragile. "Things that seem like simple requests to the public or leadership take us a long time due to our technology limitations," one SME noted.

Finding/Corrective Action: CDPH should continue investing in its technology modernization efforts currently underway, including the Future Disease Surveillance System (FDSS) project. (*ID: Data and Reporting 10*)

See also the related finding *Enterprise Tech — 5* in the Enterprise Technology chapter in this AAR.

8. CDPH established a dedicated data “clean team” to manually clean, de-duplicate, and pre-process test data.

Early in the response CDPH established a “clean team” to manually de-duplicate and pre-process COVID-19 test results. This team de-duplicated data, merged and matched records, and investigated anomalies in test and case counts. Such anomalies were often inadvertently caused by LHJs and other CalREDIE users performing their own data cleaning without alerting CDPH. Sometimes these actions, such as bulk uploads or deletions, would result in negative case counts



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

or negative deaths, which CDPH would then investigate and rectify. While many data processes were automated over time, the clean team's work remained largely manual due to a variety of factors including CDPH's lack of an enterprise-level master patient index (MPI), which serves as a unique identifier across disparate systems.

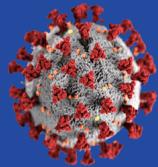
Finding/Corrective Action: CDPH should consider implementing an enterprise-level master patient index to reduce the need for manual data cleaning. (*ID: Data and Reporting 11*)

9. The emphasis on reporting perfect data detracted from other public health surveillance and analysis, delayed decision-making, and delayed the release of some data.

California's strong emphasis on data reporting sometimes hindered prompt decision-making. The pursuit of perfect and consistent data sometimes caused a delay in crucial early decisions and prevented the release of data; public criticism of potential data discrepancies impacted trust in public health. In the effort to align metrics, CDPH teams devoted significant time to investigating discrepancies that were insignificant from a public health and epidemiological perspective. This was due in part to the daily, seven days a week reporting cadence for COVID-19 test results and case rates, which was unprecedented. Many leaders and staff agreed that the State was so focused on reporting "perfect data" that it overshadowed more significant surveillance activities and delayed decision-making. Additionally, it also had "a chilling effect" on the release of some data, others noted. For instance, while CDPH began receiving requests to release case and death data by race, ethnicity, and location in Summer 2020, approval processes and discussions delayed its release by months. In the future, it will be important to "not let the perfect be the enemy of the good," in the words of one SME, and to set expectations with the public and policymakers early on regarding data reporting and any potential limitations.

Finding/Corrective Action: CDPH should establish more sustainable reporting cadences and set data expectations early during future pandemic responses. (*ID: Data and Reporting 12*)

Finding/Corrective Action: CDPH should consider establishing data and reporting confidence intervals and thresholds to expedite data release and decision-making. (*ID: Data and Reporting 13*)



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

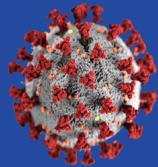
10. CDPH's data and reporting communications were driven by daily requests from the public, the media, and leadership, creating a reactive dynamic that initially hindered consistent, confident public health messaging.

CDPH initially fell into a reactive pattern of responding to external requests for COVID-19 information. As a result, data was not always communicated or presented in the most effective manner. CDPH data and reporting teams were also asked to present data to illustrate certain requested points, rather than letting the data "tell the story," according to some SMEs. Responding to questions and requests often diverted effort from important larger initiatives. Overwhelmed by a plethora of data streams and metrics and caught in a pattern of reacting to requests instead of building out a data framework, key public health messages about the data were often obscured. In hindsight, teams acknowledge the missed opportunity to develop a more proactive data and reporting communications strategy, which could have preempted potential requests, and contributed to more consistent, confident public health messaging.

Finding/Corrective Action: CDPH can proactively develop a public health emergency data strategy, and subsequent consistent messaging around public health data to set public expectations around data change and variability. (*ID: Data and Reporting 14*)

11. To make raw data available to the public rapidly, the State published COVID-19 data on its open data portal without thoroughly de-identifying it.

In early summer 2020, GovOps and State leadership directed CDPH to publish de-identified COVID-19 case, death, and hospitalization data on CalHHS's open data portal, using "small cell suppression" to further protect privacy. In June 2020, while CDPH teams were in the middle of suppressing the data in accordance with de-identification guidelines, CDT published the unsuppressed data on the State's open data portal. The unsuppressed data remained public until August 2020, when CDPH assumed ownership of the properly suppressed data on the CalHHS open data portal. The experience highlighted the need for earlier involvement and better communication between response partners working on public data reporting.



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Finding/Corrective Action: In the future, CDPH and its response partners should develop policies across departments, and coordinate to ensure that data is properly suppressed and de-identified according to legal and ethical constraints prior to publication. (*ID: Data and Reporting 15*)

12. Early in the response, data reporting functions were decentralized, leading to duplication of effort, data inconsistencies, and a lack of coordination.

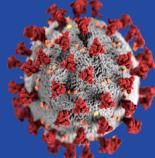
When the pandemic response began, CDPH's data teams and epidemiology teams worked together within the Center for Infectious Diseases (CID). However, in April 2020, a separate team called "Lane 6" that reported collectively to CDPH, CalHHS, and State leadership was created. This created a silo, and the Lane 6 team ended up duplicating existing data reporting work. Several months later, CDPH established additional data teams to support county data monitoring without clear coordination or reporting structures. As a result, there were frequent internal inconsistencies in case rate calculations and metrics. Urgent requests for information exacerbated the confusion, resulting in data errors and inconsistent numbers. Leadership was also initially unaware of certain data streams, creating further challenges in answering data questions. Both leadership and staff acknowledged that the initial execution to decentralize data missed early opportunities to standardize and streamline data.

Finding/Corrective Action: CDPH should follow ICS structures during emergencies to ensure consistent, streamlined communication and avoid creating multiple, decentralized data teams in an emergency. (*ID: Data and Reporting 16*)

Finding/Corrective Action: CDPH should designate a Chief Data Officer (or equivalent) and/or a Data Branch in any emergency response to oversee and coordinate all data streams. (*ID: Data and Reporting 17*)

13. To generate statewide reports CDPH relied on thousands of external stakeholders to submit timely and accurate COVID-19 data. This dependency was challenging, as CDPH was responsible to report daily on data that it did not control.

While CDPH was responsible for most COVID-19 data reporting, it was an aggregator for much of this information, which is submitted by labs, providers, LHJs, and hospitals. This includes separate data from 2 large



CDPH COVID-19 After Action Report

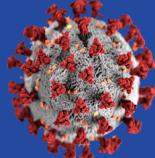
Chapter 19 – Data and Reporting

LHJs comprising almost 1/3 of the state population. As a result, CDPH teams fielded questions about data that it had little control over and were often held responsible for the quality and completeness of the data. “It’s really challenging that we need data from so many partners, and those connections weren’t well established at the beginning,” one SME noted. Even when mandates were changed, for instance, requiring labs to report race and ethnicity information with COVID-19 test results, third-party data systems were not always able to collect this information. To improve the quality of test data in 2020, CDPH collaborated extensively with labs to help them configure their data systems appropriately. To improve the quality of vaccination data in 2021, the State contracted a Third-Party Administrator and consultant Data Strike Team to conduct outreach to providers and help them report their immunization data properly. Still, many agreed that CDPH needed greater enforcement authority, enabled by legislation, which would compel third parties to improve their data submittals and report more consistent, standardized data to the State.

Finding/Corrective Action: CDPH should explore and recommend legislative solutions that would help improve the quality and consistency of data submitted by third parties. (*ID: Data and Reporting 18*)

14. Early in the response CDPH rotated staff in the data teams, but abandoned this approach several months later after determining that it is not suited to the highly technical work of data and reporting.

During the early stages of the pandemic response, CDPH attempted to establish a traditional emergency response structure based on rotating staff for data and reporting functions. However, this approach proved unsustainable for the data and reporting teams. The constant turnover associated with rotations hindered training efforts, as new staff would complete training just as their rotation ended. It became evident that the technical nature of data and reporting work required longer periods to develop essential institutional knowledge and technical expertise. Consequently, in the summer of 2020, CDPH shifted its focus to hiring and redirecting personnel for long-term involvement in the COVID-19 response. This change improved staffing continuity and eliminated disruptions caused by rotations. The experience taught CDPH that while rotational approaches may be suitable for certain aspects of public



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

health emergency response, they are not well-suited for intricate public health investigation or data and reporting tasks.

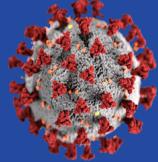
Finding/Corrective Action: In the future, CDPH should establish much longer rotations (e.g. several months) for technical positions to balance staff burnout with time for orientation and familiarity with the roles and identify key cross training that may be needed across the department to improve the response. (*ID: Data and Reporting 19*)

15. CDPH lacked adequate skilled technical and informatics staff to meet the demands of COVID-19 data reporting.

Early in the pandemic response, CDPH faced a shortage of technical staff skilled in data reporting, modeling, informatics, analysis, and management. Historically CDPH has a deep bench in clinicians and epidemiologists but lacked the technical staff to meet the overwhelming need for COVID-19 data. As a result, CDPH struggled with data accuracy, timeliness, and centralization. Technical teams were overwhelmed as COVID-19 data and reporting volumes and responsibilities increased rapidly. The scrutiny associated with daily reporting requirements further contributed to burnout and low morale amongst these teams. Additionally, although CDPH began hiring more technical staff in summer 2020, the work of recruiting, training, and onboarding staff often fell to data and reporting teams, which detracted from their operational response work. The pandemic demonstrated the centrality of data to an effective public health response, and the importance of hiring, retaining, and training staff who understand its complexities. This will become increasingly important for other current public health data and response activities as data modernization occurs.

Finding/Corrective Action: CDPH should further invest in its data and reporting workforce by identifying, recruiting, and hiring employees with the necessary technical and informatics skills. (*ID: Data and Reporting 20*)

Finding/Corrective Action: CDPH's administrative support staff, and not data and reporting team members, should manage the recruitment and onboarding of new technical staff. (*ID: Data and Reporting 21*)



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

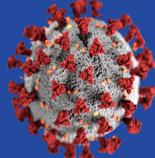
16. The lack of adherence to a formal Incident Command Structure created communication challenges between data and reporting teams and CDPH leadership.

The communication between data and reporting teams and CDPH leadership was initially disorganized and problematic due to the lack of adherence to the formal Incident Command System (ICS) and the lack of emergency preparedness training. Many were unaware that regular duties and reporting chains can differ in an emergency response setting. Although an ICS structure was in place, CDPH did not consistently adhere to it, and requests for information from CDPH leadership were made through various ad hoc channels such as email, Slack, phone calls, and text messages. This lack of a consistent reporting and communication structure resulted in duplicate requests and requests being sent to incorrect teams. Additionally, it was challenging for the teams to keep track of information requests and determine which ones should take priority. Data and reporting teams also struggled to communicate effectively with certain task forces, which were located outside of the traditional response structure.

Finding/Corrective Action: CDPH should follow established ICS communication structures, establish consistent ICS structures and training/orientation for complex responses, develop a formal process for internal information requests, and avoid using informal channels for significant requests during emergency responses. (*ID: Data and Reporting 22*)

17. CDPH lacks a real-time, consistent source of hospitalization data, which limits hospital surveillance activities.

At the start of the pandemic, the State had no insight into the status of California's hospitals, making it difficult to determine when hospitals were reaching their capacity. In March 2020, a federal hospital survey was established in conjunction with the California Hospital Association, which collected data points and metrics established by the U.S. Health and Human Services agency. However, this daily survey provided aggregated, facility-level data, which proved challenging to work with. In July 2021, CDPH issued California-specific guidance directing hospitals to report patient-level discharge data weekly for COVID-19 positive cases. Over time, CDPH's data teams relied more on this more detailed



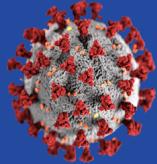
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

dataset than the federal survey. However, CDPH still lacks access to granular real-time statewide hospitalization and surveillance data.

Finding/Corrective Action: CDPH should explore legislative and technology solutions to improve collection of ongoing hospitalization surveillance and utilization data, potentially through anonymized, secure sharing of Electronic Health Record (EHR) data. (*ID: Data and Reporting* 23)

DRAFT



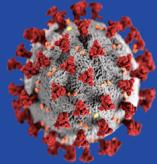
Analysis of Activities

This section elaborates and provides more detail on the findings, corrective actions, and lessons learned that are presented in the Main Strengths and Successes and the Main Challenges and Lessons Learned sections.

Data Teams and Workstreams

Initial Creation of Multiple, Different Data Teams within CDPH

- At the start of the pandemic in early 2020, CDPH data and epidemiology staff assigned to the early COVID-19 response worked in the same section. The “data team,” as it was known, was comprised of data managers and data informaticians, who worked alongside the “epi team,” which was made up epidemiologists. While the data team focused on data infrastructure, management, and informatics, the epi team worked with the data to conduct public health surveillance, interpretation, and analysis. Both teams worked closely together and were located within CDPH’s Center for Infectious Diseases (CID). Within CID, they worked within the Division of Communicable Disease Control’s (DCDC) Immunization Branch (IZB). Since IZB housed other respiratory viruses such as influenza, early on a decision was made that COVID-19 work would be housed within this branch. During these early days, the data and epi teams working on COVID-19 focused on case data, which was derived from local public health reports and laboratory test results. (When vaccination data became available in December 2020, the Vaccine Task Force’s Reporting and Analytics team handled this data stream.)
- However, in late April 2020, a group of epidemiologists and data managers were moved to a newly created separate team known as “Lane 6” that reported directly to the CDPH Directorate and CalHHS leadership. This team’s direct link to leadership enabled it to access resources more quickly. However, Lane 6 was tasked with very similar activities that were already underway and ended up duplicating much of the work that had already been started by other data and epi teams in slightly different ways. Lane 6 began working on data modeling, but was separated from the rest of the response effort and was not in a position to collaborate with the data, reporting, and epidemiology teams.
- Consequently, soon thereafter staff discovered the existence of “rogue scripts” calculating different metrics, leading to reporting inconsistencies.

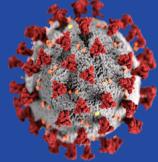


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

In the following months, additional data teams were established, including the Blueprint for a Safer Economy team and the County Data Monitoring team. These teams were often a surprise to staff, and it was not clear who the new teams reported to or what their responsibilities were. “There was no coordination of all these teams that were set up to use the main source of data,” one SME explained.

- With so many teams using the same test data to calculate COVID-19 case rates, internal inconsistencies were common. A simple example of this was using different denominators to calculate case rates. For instance, some teams would include incarcerated populations in their calculations, while others would not, which would lead to different test positivity rates. With so many teams using the test data in different ways, there would be many differences in case rate calculations and metrics. As a result, “we’d spend most of our days investigating why there are differences,” one SME recalled.
- Often, urgent requests for information from CDPH leadership that were made to multiple teams exacerbated the confusion. When one team could not produce the answer quickly enough, leadership would ask another team. Sometimes, these requests would reach inappropriate specialists, for instance, to technical SAS coders who lacked context of the request and would provide the wrong numbers. Ultimately, the fact that so many different analysts were making edits to the data at different times contributed to inconsistent numbers, often caused by errors in code and manual typos.
- The fragmentation of data was also related to CDPH’s initial lack of an Incident Command System (ICS) structure. Instead of following ICS reporting chains, requests from leadership were sometimes based on existing relationships. Various leaders were requesting similar data from different individuals and teams. This led to duplication of effort and further discrepancies in calculations.
- CDPH leadership were also surprised by the discovery of new data stream teams. According to one, “there were data collections activities going on that we didn’t even know about.” Leadership and staff agreed that the initial decentralization of data created missteps and that early opportunities to standardize and streamline data needed to be prioritized. For additional discussion of Lane 6, see the Epidemiology and Surveillance chapter in this AAR.



CDPH COVID-19 After Action Report

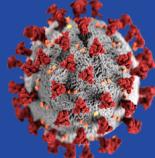
Chapter 19 – Data and Reporting

Centralized Data Processing and Informatics Section Formally Established

- Towards the end of the summer and early fall 2020, there was a realization that case data needed to be better centralized and standardized. According to leadership, a tremendous amount of thought and energy went into synchronizing the data streams. “We had to establish common ways of talking about data and use the same numbers, otherwise it's all about tracking down why the numbers are different,” one noted.
- Lane 6 was discontinued, and in late fall 2020, CDPH created a new Data Processing and Informatics Section (DPI) to assume responsibility for all case data processing. DPI was housed within the Coronavirus Science Branch, which was later renamed to Disease Informatics Surveillance Clinical and Outbreaks (DISCO) team. The Science Branch originally resided outside of CID/DCDC. However, most of the staff were pulled from CID and DCDC, including epidemiologists from the Immunization Branch, Sexually Transmitted Diseases Control Branch, Office of AIDS, and others.
- DPI was comprised of informatics specialists, epidemiologists, data analysts, and programmers who receive, process, and prepare raw data prior to analysis and interpretation by CDPH epidemiologists. This reorganization streamlined and centralized data operations and processes. As additional data collection was initiated and new data streams were established, DPI collaborated across all COVID-19 response teams and programs to merge, centralize, and standard datasets. CDPH also maintains a CalREDIE program comprised of coders, informaticians, and data managers. Since this team works with the technical details of incoming laboratory test result data, they collaborate closely with DPI.
- For future emergency responses, SMEs recommended that CDPH maintain a centralized data and informatics section tasked with centralizing and managing all data utilization and integration requests. Additionally, SMEs noted the need for a Response Chief Data Officer (or similar role) who has visibility into all data streams and is empowered to make integration decisions. This could be accomplished by updating the MHCC ICS organizational chart to include a Data and Informatics Branch or Section.

Data Reporting

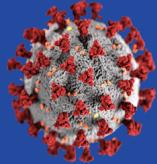
New Reporting Requirements Included Increased Reporting Frequency and New Data Points



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

- As COVID-19 quickly escalated in early 2020, so did the demands for data reporting. Prior to COVID-19, CDPH was accustomed to reporting weekly to the CDC and updating its website annually. This cadence provided ample time for data validation, deduplication, accessibility, and visualization considerations. During prior emergency responses, such as the measles outbreak of 2014-2015, CDPH had provided daily or weekly case count updates, but volumes were low. However, in early 2020, CDPH was suddenly under pressure to meet increasing demands for COVID-19 information from the State, the CDC, and the public. These new reporting expectations and requirements were unprecedented in terms of frequency, volume, and duration. CDPH quickly began reporting COVID-19 case data 7 days per week, which was unprecedented in a public health response, according to one SME. In comparison, during flu seasons CDPH typically reports on a weekly basis for limited time periods. From an epidemiological perspective, weekly reporting is the accepted standard and daily reporting is not usually epidemiologically relevant with slower moving or lower morbidity diseases. For the purposes of routine public health surveillance, epidemiologists are interested in 7-day averages since these reveal more important disease trends than day-to-day changes.
- During this pandemic, doubling time (e.g., the time required for the number of infections to double) was extremely important to inform the scale and scope of the emerging outbreak. When SARS-CoV-2 detections were emerging and expanding quickly, doubling time helped understand the emerging impact on communities and the healthcare system.
- Thus, CDPH's epidemiology and data teams began reporting case and other data 7 days per week. These new, unprecedented reporting requirements were accompanied by immense pressure to ensure presented data was timely, accurate, and addressed the increasingly numerous and varied questions stakeholders wanted answered to inform their surge planning and response.
- For instance, while reporting initially focused on case numbers, reporting efforts quickly expanded in March 2020 to encompass testing capacity since the early lack of adequate testing was a huge barrier to adequate detection and public health response. CDPH teams started receiving questions from leadership about how much COVID-19 testing was occurring in the State. However, there was no clear answer to this question since labs had never been required to report negative test



CDPH COVID-19 After Action Report

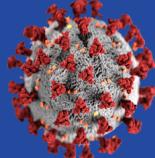
Chapter 19 – Data and Reporting

results. With only positive test results being reported, there was no clear way to understand testing volumes in the state. Starting in March 2020, CDPH required laboratories to report negative test results and provided guidance on how to do so. “It was new for negative tests to be reported to public health,” one SME noted, and there was initially no clear way to collect the data and provide the requested reports.

- However, at that time, the CalREDIE system was not yet able to accept testing capacity data. As a temporary solution to bridge this gap, external contractors built a cloud-based reporting portal that was used by labs to capture their testing estimates. This system, the Lab Testing Metrics Application (LTM) went live in early May 2020. It was set up to receive aggregated, de-identified test data that data teams consolidated into an Excel-based dashboard with other reports that distributed to CDPH, Cal HHS, and other State leadership. LTM was a “necessary evil,” according to SMEs; while it met a need to report on daily testing capacity, it was based on estimates rather than actuals. Data integrity was also a concern, since the State was relying on a mix of datasets that included actuals and estimates. Once the precedent to use testing estimates had been established, it was hard to “wean off” of it, according to interviewees. However, the historical under-investment in electronic laboratory reporting (ELR) left the State with no other options. In early December 2020, CalREDIE’s ELR functions were implemented and the use of LTM to report on testing capacity were discontinued.
- The new requirement to report all test results—both positive and negative—established a precedent that many SMEs felt is permanent. For instance, during the mpox outbreak in 2022, CDPH required reporting both positive and negative test results. Monitoring test positivity (which was emerging as a way to monitor trends in other respiratory virus surveillance prior to the pandemic) emerged as a new capacity developed during this response that can now be used for other disease surveillance.

Federal Reporting Cadences were Established and Standardized

- Federal reporting of California’s COVID-19 case data took multiples forms. In Spring 2020, in response to CDC’s request, California and other States began reporting COVID-19 line list case data to CDC’s Data Collation and Integration for Public Health Event Response platform (DCIPHER). Line list (also known as line level) data contains key information about each case in an outbreak, with each row representing a case and each column



CDPH COVID-19 After Action Report

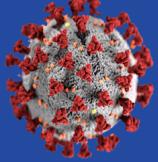
Chapter 19 – Data and Reporting

representing a variable such as demographic, clinical, and epidemiologic information. In March 2020, in addition to reporting line level data, CDPH also started reporting public health (commercial testing was not yet available) laboratory data via DCIPHER in response to the CDC's request. Additionally, once aggregate case counts were made available to the public on California's open data portal, this summary data was also reported to CDC via API. Lastly, CDC's partnership with the American Public Health Laboratory (APHL) initiated another federal reporting requirement, as CDPH also began reporting ELR data formatted to APHL standards. This was known as CELR Reporting (COVID-19 Electronic Laboratory Reporting).

- One of the earliest challenges with meeting these federal reporting requirements was generating a Person Under Investigation (PUI) number. In early spring 2020, testing for COVID-19 was only being conducted by the CDC and individuals had to meet requirements to be tested. CDC required everybody who was being tested to obtain a PUI number. Early on, LHJs and providers needed to call CDC directly to obtain this number. However, soon thereafter CDC began requiring the State to generate PUI numbers. To enable this, CDPH's informaticians built an application where LHJs could log in and initiate a request to generate the PUI number. As soon as commercial testing became available in early summer 2020, this process became obsolete. According to one SME, "it was a lot of effort to make it happen, it lived for a month, and it was not so useful."
- Over the course of the pandemic, the CDC's case data reporting requirements changed. At certain times, CDPH only reported confirmed cases, but other times, it had to report confirmed and probable cases. In response to CDC requests, CDPH worked with APHL to automate reports, which included attempting to establish a "pull" mechanism to obtain case data in near-real-time. However, it proved too difficult to automate CELR data, and when reporting requirements were decreased to twice per week, this effort was abandoned.
- Most federal reports that were established early in the pandemic continue today and are still in place as of June 2023.

CDPH, CalHHS, and State Leadership Established Reporting Cadences, including Unprecedented Daily Reporting

- Early State data reporting efforts in Spring 2020 were focused primarily on case reporting. When State reporting first began, CDPH data and

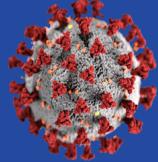


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

epidemiology teams would email a PowerPoint report every night, 7 days per week to CDPH leadership. An abbreviated version of this report would go to CalHHS State leadership. The PowerPoint report involved much manual manipulation and incorporated data from multiple teams on a variety of topics (e.g., test data, case data, hospitalization data, SNF data, death data). In the beginning, it took 4-5 hours to develop this daily report, and, according to one SME, “if it went out by 11 pm it was impressive.” This was due to the sequential process, in which teams had to wait for others to finish their work before they could start. With so many individuals waiting for their turn to make edits to the report, small discrepancies in the report created long delays. These discrepancies were usually numbers that were off by one digit (either due to typos or other reasons). Identifying and fixing these errors resulted in delays and prevented downstream teams from beginning their work.

- In response to requests for more data points (for instance, data about cases at non-CDPH facilities), CDPH began generating and incorporating additional data streams into this report. In the summer 2020, the slide deck ballooned to 90 slides that were updated on a daily basis. This master slidedeck was an intermediary tool that combined and synthesized multiple data streams. Leadership drew on it to create other reports that were tailored to different audiences, or converted into other formats that leaders could use to make policy decisions.
- In addition to the challenge of handling so many data streams, CDPH data teams were overwhelmed by unprecedented demand to report daily, which is unnecessary for routine disease surveillance. From some epidemiologists’ perspectives, day-to-day changes in case numbers seemed insignificant, and “we don’t need to be investigating small changes every day,” one SME noted. However, due to the rapid spread of disease, its impact on healthcare systems, and the high-stakes policy decisions and interventions required to contain the outbreak, daily reports were needed to manage the fast-moving response. In retrospect, the importance and use of this data was not communicated effectively to the front line teams working on the data, and due to the highly public and political nature of the pandemic response, it took longer than usual to decrease the frequency of reporting compared to prior infectious disease emergency responses. SMEs noted that public health epidemiologists were not engaged enough to understand the importance and utilization of their work.

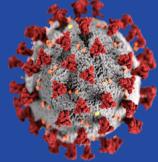


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

COVID-19 Data Reporting Subjected to Intense Interest and Scrutiny

- In addition to facing the unprecedented and unexpected demand for daily reporting of extremely high volumes of data, CDPH also faced intense scrutiny regarding COVID-19 data accuracy, especially case counts. This scrutiny came from the media, the public, lawmakers, and legislators. Many news organizations created their own data dashboards, although it was not clear what sources they were using. Citizens were also referring to the data dashboards run by CDC and Johns Hopkins University and comparing these numbers to those being reported by the State. CDPH was caught off guard by the scrutiny and demand for data and statistics. “Expectations changed, and people wanted more visibility into everything,” one leader noted.
- Consequently, CDPH fell into a pattern of reacting and responding to public and political scrutiny regarding its data, which often involved explaining numerical discrepancies—even if they were insignificant. If reported numbers were off by just a single digit, CDPH teams would spend hours attempting to find the reason, even though the answer would not result in a change in policy or response. “It doesn’t change your response if your number is 290 or 291,” one leader noted. Due to the impact of discrepancies on credibility of public health departments and CDPH, CDPH leadership and data teams devoted considerable energy to this level of granularity, starting in March 2020 and continuing through 2021.
- There was a myriad of reasons that led to reporting discrepancies. At the start of the pandemic, CDPH’s legacy technology systems became unstable due to the volume of data these systems were processing. Additionally, CDPH relied on labs and health providers to input the data, and many times these third parties did so incorrectly. Other times, LHDs would make adjustments in CalREDIE that resulted in numerical changes. For instance, at one point in the pandemic response San Diego County reclassified over 50 deaths as non-COVID-related, which retroactively impacted the State’s COVID-19 death rate. Since CDPH was reporting daily, the discrepancy was immediately visible, and State teams had to devote hours trying to figure out why the day’s death count was negative.
- Ultimately, leadership and staff agreed that it was not a constructive use of time to track and report data with this degree of granularity. “I hope in the future we don’t spend so much energy hunting down insignificant details,” one leader noted. Beginning in 2022 more tolerance was allowed



CDPH COVID-19 After Action Report

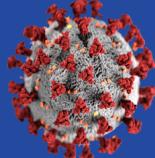
Chapter 19 – Data and Reporting

in case, test, and death data and the pressure to report perfect data waned.

- The intense interest and scrutiny into CDPH's COVID-19 data reporting also resulted in a fear of inaccuracy and a reluctance to release some data. These negative consequences were in many ways the result of a "hyperfocus on real-time data," according to one SME. Rather than acknowledging the fallibility of COVID-19 data and systems, "we just decided not to release data because we were concerned those imperfections would be visible," another SME added. In one example, some datasets individuals of Filipino descent were miscategorized as native Hawaiian Pacific Islander or Asian. Instead of releasing this data with a qualifier, CDPH temporarily did not release it at all.
- SMEs noted that an important lesson learned for future pandemic response is to proactively manage public expectations about data reporting. This could include messaging around margins of errors, data variability and limitations, and tolerance levels. This would help temper expectations and maintain trust regarding data reporting and free CDPH to focus on the data needed for policy decision-making.

Early, Significant Data Reporting Decisions Impacted Death Data and Reporting Dates

- In early 2020, as new data streams and teams were being established to meet new reporting requirements and requests, two significant decisions were made related to data sources. The first was the decision to use Cal OES death collected data instead of CDPH death data, and the second was the decision to report COVID-19 cases by what is known as report date rather than episode date. The technical nuances of these decisions were not immediately visible, but impacted the State's data metrics and data quality that became clear over time.
- Death surveillance data for COVID-19 is kept in CalREDIE along with other disease surveillance information. Early in the response, CDPH began providing guidance to LHJs to help them record death data in a standardized fashion and so that COVID-19 associated deaths were reported accurately. This involved a critical distinction between dying "from" COVID-19 and dying "with" COVID-19. In other words, distinguishing whether the cause of death was directly attributable to COVID-19, or whether the death was caused by something else (e.g., a

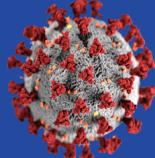


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

traumatic head injury) while the patient also happened to have COVID-19.

- However, in early Spring 2020, State leadership tasked the California Department of Technology (CDT) with collecting COVID-19 data from agencies, including death data. As a result, a key decision was made to report death data not from CDPH and the State's disease surveillance system, but from Cal OES's mass fatality coordination program, since this program has established relationships with coroners.
- Unfortunately, this alternative source of death data, which was used and publicly reported for over a year, lacked the rigor and accuracy associated with epidemiology and public health surveillance. This data was a line list in Excel, collected manually from county websites, and it lacked important epidemiological details, including age, name, date, and laboratory confirmation. CalREDIE, in contrast, possessed much more comprehensive and reconciled COVID-19 death data.
- “Someone made a decision to use another source of death data, not public health surveillance data,” according to one SME. CDPH data teams, concerned with data integrity and accuracy, spent a year working to reverse this decision and “convincing everybody up the chain that this was not the source of information we should be using,” another SME noted. Eventually, CDPH SMEs were granted access to the Cal OES death data in July 2020. However, it was not until spring 2021 that the source of California’s death data reporting was changed to CalREDIE. A year after, in 2022, CDPH took over full management of all COVID-19 data from CDT.
- Traditionally, disease conditions in CalREDIE are reported by episode date, which reflect the onset of illness. However, decisions outside of CDPH at both state and federal levels were also made to report out COVID-19 case data by report date rather than episode date. The episode date is electronically calculated in CalREDIE based on the ELR message, whereas the report date is simply the date that the case was reported. However, COVID-19 cases were initially reported by report date. Using the report date has less epidemiological relevance according to SMEs; it simply reflects “the date we heard about it rather than the date of illness,” and is subject to reporting lags and delays.
- Tabulating by report date involves an implicit lag. For instance, if 10,000 new cases were reported, it does not mean that 10,000 new people were



CDPH COVID-19 After Action Report

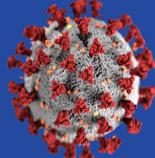
Chapter 19 – Data and Reporting

infected the day before, but rather that they were infected between the day before and 3-4 weeks ago. Reporting by episode date, in comparison, by design includes a standard 7-day lag period to account for disease onset and progression.

- However, the technical complexities and nuances of these distinctions were not initially considered in early data reporting decision making. “Decisions were coming fast and furious, people didn’t think it through, and that negatively impacted our analytics metrics,” one SME noted. Additionally, it took time for the technical subtleties of this reporting distinction to be widely understood. In 2020, data and epidemiology team members began suggesting that the case reporting be changed to reflect the episode data, but it took until spring 2021 for this change to be made. “We got stuck in this pattern, and it took a year and half to shift mindsets,” another commented.
- Finally, the State dashboard was updated to include episode date in addition to report date, which SMEs considered a major success. Making this change entailed obtaining leadership and executive approval and was a lengthy process. Attempts to make this change at the federal level were less successful. In early spring 2021, California was one of the only States to have its case metrics updated on the CDC website to reflect episode date. However, after several weeks, there was so much confusion about the lack of alignment between State and federal reports, that this change was undone.
- For future pandemics, many agreed that it was important to think through data reporting ramifications from epidemiological and public health perspectives, rather than making rapid decisions that are not informed and influenced by the public health SMEs. SMEs pointed out that this lesson had been learned for the mpox outbreak in mid-2022, with reporting based on episode date.

Automation Helped Ease Daily Reporting Burden

- Daily reporting was especially problematic in the first six months of 2020 with such high volumes prior to automating many reports. However, this grew easier over time as CDPH increased efficiencies and automation through data standardization and other improvements. This included implementing the California COVID-19 Reporting System (CCRS) and the Snowflake data warehouse in fall 2020 (which is discussed in greater detail in the Data and Technology section of this chapter). The ability to direct

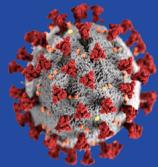


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

COVID-19 data through CCRS allowed CDPH teams and programs to access their data more quickly, which facilitated timelier reporting and processing.

- Over time, the daily slidedeck report was shortened and much of the data inputs were automated, especially in the second half of 2021. However, the slidedeck report still required several hours of manual compilation.
- While legacy tools were needed in 2020, throughout the pandemic CDPH's technology, data, and informatics teams made major advancements, including developing internalized dashboards and automated file sharing. However, there has been a reluctance from some to embrace these improvements, and some still prefer printing the dashboard reports or viewing data in a CSV file. "We did a lot of work to automate things," one SME noted, but with frequent requests for customizations, "you lose a lot of the benefit of automation."
- CDPH's data and informatics teams have redesigned data reporting processes to make them more efficient and automated, and it has been challenging when some internal stakeholders insist on receiving data in legacy formats. According to one SME, "we have the skillsets and infrastructure to deliver data in a streamlined, savvy way now." Not all CDPH staff may be aware of the more sophisticated reporting options that are now available, which presents an opportunity for organizational change management, training, and education. Many pointed out that a middle ground exists between the real-time, fluctuating dashboards that are constantly changing, and static files and paper associated with legacy reporting. CDPH's reporting tools can be configured to send out PDF snapshots of real-time data. "We can do both now," one SME added; "we can provide real-time, savvy insights as well as accommodate the desire to have something concrete."
- Following the introduction of more automated tools, another challenge emerged regarding the best use of data team resources. Informatics teams continued to receive requests from leadership for ad-hoc reports. Consequently, sophisticated informaticians who were working on longer-term solutions would frequently be diverted to handle more urgent rudimentary requests. While there is a need for informatics work, the skillset is not widely available in CDPH. Therefore, one SME noted that "we need to do a better job of understanding the charge and skillset of certain



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

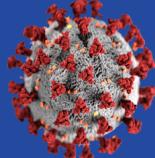
groups, and not tying up informatics' hands with things like manually compiling reports."

- At the same time, informatics and data teams acknowledged the need to be "better champions for justifying this approach," according to one SME. This would entail thinking through data and reporting needs strategically and comprehensively and anticipating future data points that may become important.

Data Submission

CDPH Worked with Labs to Receive Test Data

- Although CDPH reported on COVID-19 data, CDPH received this data from the thousands of labs, providers, LHJs, other testing sites, and hospitals who were testing and treating patients. This created a dynamic in which CDPH was held responsible for the quality and timeliness of data but had little control over it. As one leader noted, "We don't collect the data, we are just a receptacle of the data." While there was ongoing work throughout the pandemic to improve the quality and completeness of multiple data streams, there were two significant efforts that involved intensive collaboration between CDPH, who was receiving data, and the external stakeholders who were sending data. The first effort involved working with labs on test and case data in the first half of 2020, and the second project involved working with providers on vaccination data quality in early 2021.
- In early spring 2020, as testing for COVID-19 shifted from the CDC to individual laboratories, California pushed to expand testing capacity across the State. In response, more and different types of laboratories began offering COVID-19 testing, each with different reporting capabilities—for instance, an independent pop-up clinic performing COVID-19 testing would have fewer technical capabilities than a global corporation such as LabCorp or Quest Diagnostics.
- CDPH's CalREDIE program team worked with the labs to develop creative ways to receive test data into CalREDIE, the State's disease surveillance system. In early 2020, this included enabling manual lab reporting in addition to accepting Excel/CSV files while working to onboard labs to the ELR functionality. As mentioned previously, beginning in March 2020 labs were required to report both negative and positive test results.

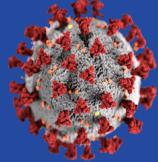


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

However, the capability to submit and receive negative test results was not a functionality that could simply be turned on.

- Whenever CDPH asks labs to send additional or different data that they were not previously submitting, labs must configure their individual laboratory information management technology systems to comply. Typically, when new lab reporting requirements are instituted, the CalREDIE program team communicates and coordinates with the labs to ensure that the data being submitted follows appropriate standards and can be processed by CalREDIE. In early spring 2020, with so many new labs signing up to offer testing and meet new reporting requirements, labs struggled to report accurately and timely data.
- In addition to continuously onboarding new labs and helping them establish electronic reporting capabilities, the CalREDIE program team worked with labs who were reporting incorrectly. This was sometimes a tedious process, as the root causes of the incorrect data varied widely. As the pandemic progressed, different types of COVID-19 tests became available, and rapid increases of hundreds of test providers evolved. While early testing was primarily PCR-based, the subsequent arrival and proliferation of antigen testing in late 2020 and early 2021 prompted new test reporting requirements that required labs to configure their systems and processes to meet. Initially, the CalREDIE program team devoted to these issues only had 3 people, but due to the complexity of the testing landscape and the number of labs it had to support, this team grew to 15 by 2022.
- CDPH worked intensively to educate and collaborate with labs to improve the quality of their data submissions. However, some SMEs felt that CDPH needed greater legislatively-mandated licensing and/or enforcement authority to require labs to comply with reporting requirements. Empowering the State with this authority would improve data quality and completeness. “If we had that power, data would be easy to get and easy to submit,” one SME noted.
- Beginning in mid-2020 CDPH also launched multiple projects to improve its technology systems, including CalREDIE, to enhance data reporting. Teams agreed that improving the way that data was received and processed was the most critical intervention. According to one SME, “you can do all the provider outreach in the world, but if we don’t have the technology infrastructure to analyze and interact with the data, we are



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

hamstrung.” Over the course of the pandemic, the CalREDIE team worked with its developer to change the way the system received data. The implementation of other systems and tools, including the COVID-19 Reporting System and the Snowflake data warehousing tool, also improved data reporting. For further discussion, see the Data and Technology section in this chapter.

Data Strike Team Helped Providers Improve Vaccination Data Quality

- In early 2021, several weeks into the State's COVID-19 vaccination program, it became clear that vaccination administration data was inaccurate. This was due to several factors, including that it was challenging for providers to meet the mandatory reporting requirements associated with the program. Some providers failed to report their data in a timely manner, others lacked the appropriate system accounts, and others used incorrect IDs. The State brought in a Third-Party Administrator (TPA) to help with the vaccination campaign, and one of the TPA's first initiatives addressed data quality issues with a consultant Data Strike Team. The team called thousands of providers, tracked down those who were not reporting, and helped them create accounts and use the proper IDs to report their vaccination data. For further discussion, see the Vaccines chapter in this AAR.

Third-Party Vendor Systems Have Limitations to Report Race and Ethnicity Data

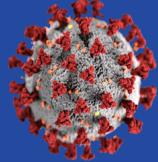
- In early 2020, when patients went to labs to get tested for COVID-19, the result often did not include detailed demographic information, such as race and ethnicity. Providers who ordered lab tests had not been historically required to include this information in the test order. However, this changed in Summer 2020, when labs were mandated to report race and ethnicity data along with their COVID-19 test results. In March 2020, the CDC issued [federal guidance](#) (effective April 2020) that required all labs to report detailed demographic information (including age, sex, race, and ethnicity) along with COVID-19 test results. Similarly, CDPH [issued guidance on July 28, 2020](#), that expanded data reporting to include these detailed demographics.
- Many labs struggled to comply with this new mandate, in part because their lab reporting systems did not possess the functionality to capture this information. Especially for auto-processed lab results, CDPH would receive

poor quality demographic data, with these fields marked “unknown.” Furthermore, many lab systems are not set up to accommodate more than the five major race and ethnicity categories. These are referred to as the “OMB” categories since they are designated by the federal Office of Management and Budget Standards. The OMB categories have one category for ethnicity, and five minimum categories for race (Black or African-American, Asian, White, American Indian, or Alaska Native, and Native Hawaiian or Other Pacific Islander).

- These limitations are not unique to lab testing data systems, but also apply to provider immunization data systems. If these third-party vendor systems only report high level race and ethnicity categories, then “we can’t disaggregate the data and get into community-level race data analysis,” according to one SME. The inability to collect more granular race and ethnicity data hinders effective policymaking to address inequities in public health outcomes. When large populations are bundled into broad data categories, it is not possible to discern different outcomes for sub-populations, and “you could be missing groups that are being left behind,” one SME explained. Public health officials need detailed race and ethnicity data to help understand the root causes of inequities before implementing policies and sharing with partners and communities to address them.
- CDPH data and reporting teams identified workarounds, including the use of other datasets that contain more complete race and ethnicity information, for instance birth and death records. Oftentimes, death certificates contain more comprehensive race and ethnicity data, and so these datasets can be integrated with the case registry to conduct more detailed analysis about subgroups.
- In January 2023, [AB 1797](#) went into effect, which required California healthcare providers who administer vaccines to report race and ethnicity information for each patient into CAIR2. Additionally, SMEs suggested that legislation be considered to help increase the consistency of race and ethnicity data reported through individual provider and vendor systems.

Two Large LHJs Do Not Use the Statewide Disease Surveillance System

- The fact that two of California’s largest LHJs, Los Angeles County and San Diego County, do not use CalREDIE for disease surveillance was



CDPH COVID-19 After Action Report

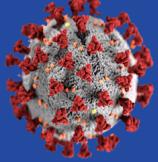
Chapter 19 – Data and Reporting

problematic for CDPH's data and reporting teams. The total population of these two counties makes up more than 33% of California's population. Instead of using CalREDIE to report COVID-19 test and case data, these counties send data to CDPH in separate files, which were notably extremely large during peaks of cases reported. CDPH then combines this data with the processed test and case data from CalREDIE to create the official COVID-19 case registry, a dataset that serves as the source of truth for Statewide case reporting. Some SMEs expressed frustration with this dynamic.

Data Quality and Remediation

“Clean Team” Established to Improve Test Data Quality

- In addition to both CDPH and Optum working with labs to help them submit more accurate test data, CDPH also established a separate “clean team” devoted to improving data quality once it had been received into CalREDIE. This team, comprised of data managers, analytics experts, and epidemiologists, focused on de-duplicating, researching, correcting data discrepancies, and adjusting case definitions. This team worked very closely with the CalREDIE program team, epidemiology teams, and the Data Processing Unit. The clean team also worked closely with LHJs who also regularly conducted their own data de-duplication and cleaning in CalREDIE. The clean team consisted of approximately five people at the Research Scientist 2 job classification level.
- Other data quality challenges were associated with the changing epidemiology of COVID-19 itself, which evolved over the course of the pandemic. Early in 2020, CDC established the interim case definition for a COVID-19 case. Typically, the CDC implemented the definition that is established annually by the Council of State and Territorial Epidemiologists (CSTE). A CSTE case definition is a set of uniform criteria used to define a disease for public health surveillance and allow public health officials to classify and count cases consistently. However, as more was learned about COVID-19, the CSTE case definition and criteria was updated frequently, presenting challenges about how to categorize data that had already been collected.
- Other data quality anomalies occurred regularly. For instance, a lab may have completed a bulk data upload, or a county may have deleted many duplicates in CalREDIE, which would affect case numbers in strange



CDPH COVID-19 After Action Report

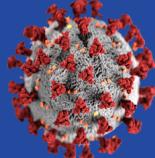
Chapter 19 – Data and Reporting

ways. In addition to monitoring and investigating anomalies, the clean team spent considerable time de-duplicating test data.

- CDPH's de-duplication work was complicated by the fact that other teams, including LHJs, were also de-duplicating records in CalREDIE, often without communication or coordination. LHJs would often perform data cleaning without letting CDPH know, and CDPH would then be surprised by unexpected changes in case numbers, such as negative case counts. Certain programs, such as the schools program, which relied primarily on antigen testing, also required intensive data cleaning. Data from testing performed at schools was fed into CalREDIE via a separate data stream and had to be manually flagged to avoid inflating case counts. Although there have been improvements in data processing, de-duplication remains a labor-intensive task.
- Along with the epidemiology and data teams, the clean team had to be prepared to answer leadership questions about test and case data. "We spent a lot of days hunting backwards, trying to figure out where things came from," one SME noted. While many of these data issues were not unique to COVID-19, they were exacerbated by the high volume of data associated with it.
- The accelerated reporting timelines without a lag contributed to data quality and completeness issues. In August 2021, CDPH teams received permission to implement a one-day lag in data quality and cleaning (but not reporting) processes. According to SMEs, it took time for leadership to realize that data cleaning and reporting did not have to occur on the same day, and implementing a lag on quality checks could have been done sooner.

Reinfections Became Difficult To Identify After Automating Positive Test Results

- For the clean team, the increasing automation of data processes had unintended consequences and introduced new challenges. In December 2020, automatic processing of positive lab test results was implemented in CalREDIE. This automation helped LHJs and CDPH better handle increasing data volumes and decreased the amount of data that needed manual review. It also helped prioritize workload. Positive test results are much more critical for LHJs to follow up on through contact tracing and case investigation, whereas negative test results are more relevant for measuring testing volumes. As one SME explained it, "test



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

volumes are an important metric for the response, but less important than case counts."

- However, there was now no way to detect reinfections of COVID-19 in the same individual. At the time, public health experts were only beginning to understand the epidemiology of reinfections, as the first U.S. case of reinfection was reported in August 2020. As it became clear that these cases were reinfections and not continuations of the initial infection, the CDC defined a "90-day reinfection rule," which meant that if an individual tested positive for COVID-19 ninety days after their first infection, it was considered a new infection. However, there was no way for CalREDIE to distinguish between original infections and reinfections.
- Consequently, when CalREDIE received a positive test result and automatically matched it, the new positive result erroneously linked to the old incident. As a result, LHJs struggled to identify cases with new positives. COVID-19 cases that were reinfections went undetected.
- By May 2021, there was a growing awareness about the prevalence of reinfections. CDPH began monitoring trends in reinfections through repeat positives in testing data. In September 2021, the CSTE case definition was implemented in CalREDIE, which included criteria for reporting new cases for individuals with previous cases. Around the same time, CDPH's systems were configured to allow multiple cases to be counted for a single person, whereas previously only one case could be counted. In Fall 2021, the CalREDIE program created an internal dashboard that allowed LHJs to review new positives attached to existing incidents. However, reporting reinfection cases still relied heavily on LHJs to manually review previous cases, create new cases, and transfer appropriate lab information.
- In Spring 2022, CalREDIE was configured to automatically detect—and count as separate—a reinfection occurring in the same individual. The creation of new fields in CalREDIE finally enabled reinfections to be tracked separately and prevented them from being conflated with old cases.

Significant Manual Data Pre-Processing and Need for a Master Patient Index

- In general, there was a significant gap in expectations between leadership's desire for quick, accurate data, and the reality of the manual effort needed to get data into a usable format. Outside of CDPH's data and epidemiology teams, very few understood the size and complexity of

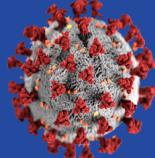
the data pipelines, and the “large pre-processing lift” required to standardize data so it can be used to make policy decisions, according to one SME. Clean team members noted the common misconception that “you could just hit a button and yield a metric.” In practice, many factors—including system limitations, lack of data standardization, reliance on data from third parties, and changing case definitions—necessitate the clean team’s manual and time-consuming remediation processes before data can be used to inform decision-making.

- One way to improve data quality, according to SMEs, would be to implement an enterprise-level master patient index (MPI)—a CDPH single unique identifier linked to an individual. This would reduce the need for data de-duplication, merging, and matching, and would enable more consistent data sharing across various program areas, including immunizations, disease surveillance, and vital records. SMEs noted that it would also promote more complete demographic data; for instance, if the immunization registry (CAIR2) contained updated race and ethnicity data on an individual but CalREDIE did not, a master patient index could be used to populate this data in CalREDIE. Implementing an MPI would need to address numerous privacy concerns, and SMEs acknowledged that it would be a complex undertaking.

Informatics and Data Integration

Data Teams were Ultimately Integrated and Centralized

- The evolution of informatics for COVID-19 progressed from rudimentary methods to innovative solutions. Early in the response, new DPI team members applied existing technology tools that had been used in the Office of AIDS, including products that helped transfer and centralize data, as well as workflows that allowed team members to collaborate on code. From March through May 2020, team members worked to standardize and streamline data and processes to enable the translation of data into more useful formats.
- For the DPI team, a major milestone was the implementation of California COVID-19 Reporting System (CCRS) and the Snowflake data warehouse in fall 2020, which helped shore up the increasingly unstable CalREDIE system. The ability to direct COVID-19 data through CCRS allowed CDPH teams and programs to access their data much earlier, which facilitated timelier reporting and processing. Similarly, the cloud-based data



CDPH COVID-19 After Action Report

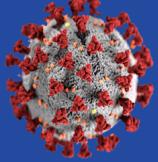
Chapter 19 – Data and Reporting

warehousing capability was a significant milestone and success. It facilitated efficient data processing, quicker dissemination to LHJs and other teams, and streamlined data pipelines for public reporting. For a more detailed discussion of CCRS and Snowflake, including the transition to the Surveillance and Public Health Information Reporting and Exchange in summer 2022, see the Data and Technology section in this chapter.

- Using CCRS, Snowflake, and supporting tools, DPI centralized multiple data streams and integrated different data streams for teams to use in their analysis. One of the most popular integrations was between COVID-19 vaccination data, housed in the State's immunization registry, and COVID-19 case data, housed in the case registry. These data sources are entirely separate and managed by different teams and task forces; vaccine data is managed by the Vaccine Task Force, which has a separate Reporting and Analytics workstream. The team collaborated with the Immunization Branch and the Vaccine Task Force to merge these datasets, so that teams could conduct various investigations—for instance, into "breakthrough" infections and reinfections in individuals already vaccinated against COVID-19.
- Leaders and experts agreed that CDPH's informatics and data integration capabilities grew from rudimentary to sophisticated over the course of the pandemic response. A key lesson learned, according to SMEs, was that this evolution should have started earlier than it did. In the future, SMEs advised engaging informatics as early as possible, as "it's never too early to start thinking about translating data from unusable to usable." The importance of informatics skills and staff demonstrated this gap in CDPH workforce that is extremely important to address for ongoing data and surveillance for routine activities and future public health emergencies.

Informatics Team Collaborated with Other Teams and LHJs to Disseminate Data

- In addition to integrating vaccination, testing, and case data, DPI also created and managed data integrations with other COVID-19 datasets, including hospitalization data, whole genome sequencing data, and supplemental data. Most supplemental data came from the schools-based testing program, which created large volumes of test data that was not in CalREDIE. School-based testing was performed primarily by third-party vendors, who each had their own testing platforms and data



CDPH COVID-19 After Action Report

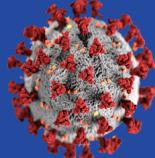
Chapter 19 – Data and Reporting

systems. These vendors would collect school testing data, aggregated, and make it available to CDPH via download or API connection. This supplemental data was used primarily to monitor outbreaks.

- By necessity, DPI worked across all COVID-19 response teams and task forces, collaborating with the owners of various datasets to centralize them. Some of these projects were more difficult than others. Hospitalization data, which is discussed further below, has been problematic since the start of the response.
- Additionally, DPI collaborated closely with legal teams responding to Public Records Act (PRA) and media requests related to data and reporting. Finally, DPI also worked closely with LHJs. In the first 6 months of the response, CDPH lacked efficient methods to distribute processed data to LHJs. “They didn’t have a great way of accessing their case data, testing data, or vaccination data,” according to one SME. Before the implementation of the Snowflake data warehouse, LHJs could download their raw case data directly from CalREDIE’s data distribution portal, but this dataset was not processed. After Snowflake was implemented, DPI granted access to LHJs so they could access their data both in raw and processed formats.

Two Hospital Data Streams Established

- CDPH relies on two primary data streams from hospitals—the hospital survey run by the California Hospital Association and U.S. Health and Human Services (CHA/HHS survey) and line list hospital discharge data mandated by [AFL 21-25](#) (issued in July 2021).
- In early 2020, there were no systems to collect important metrics from hospital facilities, such as number of occupied beds or ICU space, much less the number of hospitalized COVID-19 patients in relatively real time. CDPH called hospitals every day to compile this data for reporting purposes. However, this process was improved in March 2020 when the CHA/HHS survey was established. The CHA/HHS survey dictated a daily reporting cadence of approximately 140 metrics, including several California-specific data variables. Over time, some of these variables were fine-tuned and customized. However, the CHA/HHS survey only captures aggregated, facility-level data, which was challenging.
- CDPH data and reporting teams relied heavily on the CHA/HHS survey throughout 2020, especially during the Blueprint period in Summer 2020. According to DPI SMEs, the CHA/HHS survey was initially “very tricky,” as

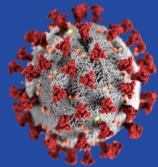


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

the team lacked visibility and control over the underlying California data. The team waited for the data to be provided in an email each day from CHA before incorporating it into its daily reporting slides and products. This data stream has since been improved.

- CDPH issued [AFL 21-25](#) in late July 2021 that asked hospitals to report line list discharge data to CDPH. This guidance directed hospitals to send line list data on a weekly basis that included all patients that had been discharged the prior week that were hospitalized and had tested positive for COVID-19. Since this data source did not stem from a federally issued survey, it is easier for CDPH to manage. When the AFL hospitalization data stream was established, CDPH SMEs worked with ITSD to establish a way to upload this data via secure file transfer. Once the approximately 90 reporting hospitals upload their data files, CDPH processes it into a single dataset and integrates it into the case registry and follow up with any hospitals that failed to report. Originally, when this data stream was established, DPI managed it but after the automation of many steps, the CAIREDIE program team took over management. As of Spring 2023, the CAIREDIE program team still receives and processes AFL line list data from hospitals weekly.
- CDPH's data and reporting teams have come to rely more on the granular AFL hospitalization data than the federal CHA/HHS survey data. The AFL line list captures data at the individual level, including the vaccination status of hospitalized individuals. This data point became very important, as it was used to illustrate that being vaccinated against COVID-19 conferred protection prevented Californians from becoming severely ill and hospitalized as a result of COVID-19.
- Ongoing surveillance of hospital capacity and function is essential to public health emergency response. While hospitalization data has improved significantly over the course of the pandemic response, there are still improvements that can be made, including clear authority and automated mechanisms and IT infrastructure to receive the hospitalization data that is useful for public health response.
- SMEs mentioned that California should explore ways to obtain improved, ongoing daily hospitalization surveillance if the CHA/HHS survey is discontinued. Specifically, the State can consider ways in which EHR (Electronic Health Record) data could be shared with CDPH's systems to conduct public health surveillance and respond to public health



CDPH COVID-19 After Action Report

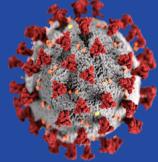
Chapter 19 – Data and Reporting

emergencies. This would require clear legal authority, and methods to keep individual health data appropriately anonymous and secure.

Data Dashboards and Visualization

Data Disseminated to the Public via Websites and Dashboards

- Beginning in November 2019, CDPH's web team began adding COVID-19 content to its departmental website. Reporting of COVID-19 facts and updates to the California public began in early March 2020. On March 4, 2020, the State released the first "Latest COVID-19 Facts" press release. This inaugurated the daily public press releases that were issued seven days per week. As discussed earlier, CDPH was not staffed for the intense, unprecedented demand from the California public and media for COVID-19 data. Some of this interest was driven by constituents who regularly viewed the Johns Hopkins Hospital dashboard (which was the first visual representation of COVID-19 data) and had expectations of similar visualizations for California's data.
- In March 2020, the California State dashboard was launched at covid19.ca.gov and it included case, death, and hospitalization data. This State dashboard was initially maintained by CDT and was designed as an easy-to-use portal for the public. CDT collected data from various State agencies to publish on the dashboard. In April 2020, CDPH changed its COVID-19 websites and dashboards to mirror the State dashboard.
- Testing data was added to the State dashboard in early summer 2020, followed by vaccination data when it became available in late December 2020. Additional programmatic websites, some of which contained dashboards, were established in 2020. These included the County Data Monitoring website in May 2020 and the Blueprint for a Safer Economy website in August 2020.
- Early in the pandemic, most (but not all) of the data disseminated on the public dashboard came from CDPH's data and epidemiology teams, yet CDPH had limited internal capacity to create and manage dashboards—thus they were managed by other redirected State departments' employees or contracted consultants. The processes by which data was added to the dashboard was unclear and confusing even to those who were participating in it. "No one really knew who did what, and there were so many different hands touching it," one SME recalled. Another noted that "data was all over the place," and that some webpages were



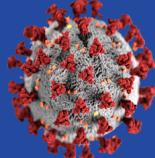
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

being updated manually based on tables sent via email. There was no single team or position that had visibility into all the official dashboards, and CDPH SMEs often had to field questions related to data that it had no control over. “We were trying to troubleshoot things we couldn’t understand,” one SME noted. For instance, CDPH teams were often asked for granular details about hospitalization data, but this data came from the federally administered CHA/HHS survey. New data would typically go live at 9:45 pm and “it was always down to the wire,” one SME noted.

COVID-19 Datasets Made Available to the Public on Open Data Portals

- Since 2014, California’s Government Operations Agency (GovOps) has sponsored data.ca.gov, a statewide open data portal created to improve collaboration, expand transparency and lead to innovation and increased effectiveness. An early adopter of open data, CalHHS developed the state’s first Agency-wide open data portal. The portal publishes data sets from the Agency’s 12 health and social services departments and offices, including CDPH. The State Portal automatically pulls datasets from the CalHHS portal.
- In early summer 2020, GovOps and state leadership instructed CDPH to include COVID-19 case, death, and hospitalization data on CalHHS’s open data portal. CDPH teams began working on de-identifying through small cell suppression, in accordance with CalHHS’s Data De-Identification Guidelines. “Small cell suppression” is an open data concept and a method used in data privacy where cells in a dataset are withheld to prevent individual identification. Even when datasets do not include any personally identifiable information, if the data is granular enough and includes characteristics like zip code, race, age, it may be possible for individuals to be identified. To prevent this, the granular data is suppressed to ensure data privacy.
- In June 2020 CDT published unsuppressed data on the State open data portal in attempt to make the data available as quickly as possible. CDPH raised concerns that this data was not adequately suppressed, but it remained posted for several months. In emergency responses, it may make sense to release unsuppressed data if the public benefit of releasing it quickly outweighs privacy considerations. As a result, CDPH soon began receiving questions from the public about the published data on the State open data portal that it could not answer.



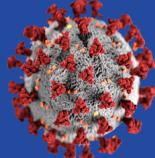
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

- In August 2020, GovOps instructed CDPH to make all data that was being publicly posted to the dashboards available on the CalHHS open data portal, including race, ethnicity, and age data. CDPH assumed management of the COVID-19 data on the CalHHS open data portal, which populated to the State open data portal. Shortly thereafter, the Snowflake data warehouse was implemented, which facilitated automated reporting to the portal. “Once the portal came back to us, and once we had Snowflake up and running, it was really smooth,” one SME noted.
- In late September 2020, CalHHS conducted a review of the public COVID-19 dashboards and concluded that data was not being suppressed properly in accordance with HIPPA guidelines. Consequently, it established processes requiring proper suppression. All COVID-19 data was suppressed on the dashboards starting in September 2020.
- In hindsight it would have been useful to have CDPH’s open data portal team involved earlier in discussions about how to present the data appropriately. In future emergency responses, CDPH teams should be included in early discussions about data suppression and publication on public portals and dashboards. This would be enabled by the use of a stronger ICS and communications structure. According to one SME, “everyone needed to be in the same room and agree on the same de-identification rules.”

Data Dashboards Redesigned and Updated

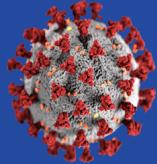
- CDPH data teams advocated for improvements and attempted to better understand data pipelines in the public dashboards. Team members began building relationships with CDT and the Office of Digital Innovation (ODI) team members who worked on the State dashboard. CDT and ODI team members lacked specific experience in epidemiology and disease surveillance data, and thus did not always understand the technical nuances of the public health data. As CDPH shared its more technical subject matter expertise with CDT and ODI, relationships were forged. “We found the right people, and they listened,” one SME noted.
- As CDPH, CDT, and ODI continued to collaborate, data sources and pipelines improved. Once it was determined that CalREDIE was the source of record for test, case, and death data, all the reports disseminated by CDPH (to the CDC, to CDPH and CalHHS leadership, to the dashboards, to the Open Data Portal, etc.) were drawn from the same source data.



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

- Several other important milestones enabled CDPH to take greater ownership and control over COVID-19 data reporting, culminating in the release of what was referred to as “Dashboard 2.0” in March 2021. These milestones included changing the source of death data from Cal OES to CDPH, which was a significant project requiring extensive collaboration. Another milestone reflected in Dashboard 2.0 included reporting cases by episode date in addition to report date (which is discussed in more detail earlier in this chapter). Lastly, California’s test positivity numbers began to be reported with both the date tests were conducted (specimen collected), in addition to when results were reported. Other major websites and dashboards, including the Open Data Portal and the Blueprint and County Data Monitoring webpages, were also updated. Ultimately, these changes in 2021 represented a significant improvement in data reporting, as they increased the data’s accuracy, precision, and epidemiological relevance.
- The State continued to make improvements, additions, and refinements to the dashboard. “Dashboard 3.0” was released in September 2021, and incorporated several significant changes. CDPH added a chart that displayed case rate by vaccination status, reflecting the integration of these two datasets. It also shifted from reporting “today’s counts” to a daily average for cases and deaths. Case rates were based on a 7-day average with a 7-day lag, and death rates were based on a 7-day average with a 21-day lag due to delays in reporting.
- In June 2022, more than two years into the pandemic response, CDPH, ODI, and CDT conducted a reassessment of public dashboards and reporting responsibilities. As a result, ownership of the State dashboard, which had previously been run by ODI, was transitioned to CDPH. SMEs noted that was another significant milestone, as CDPH gained greater control over and visibility into data pipelines and processes.
- California’s COVID-19 dashboards are seen as unequivocal successes that have evolved from nonexistence to sophistication. “The products, accessibility, and the things that are possible are pretty incredible for a State agency,” one SME noted. When fielding public questions about COVID-19 data, CDPH SMEs can often direct individuals to the appropriate dashboard with a link and screenshot. The dashboards have empowered Californians by providing answers to their questions on a “rolling basis,” with some datasets updated hourly.



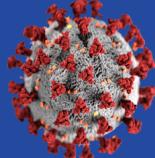
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

- Curating and maintaining the public-facing dashboards requires significant data engineering. While internal dashboards designed for public health experts (such as Snapshot, which is discussed below) often present data in a rawer form, public-facing dashboards must be curated and accompanied by annotations, caveats, and explanations. Adding these data notes is a “massive undertaking,” according to SMEs. To increase the accessibility of the public dashboards, CDPH SMEs created data dictionaries to help explain the dashboards. They agreed this was an important success to help diverse audiences understand complex data. The internal capacity for CDPH staff to create and manage dashboards is limited, and in order to maintain and develop new similar dashboards in the future, CDPH will require increased informatics and information technology CDPH staffing with the appropriate skills.

Internal Snapshot Dashboard Developed

- In addition to public-facing dashboards, CDPH also developed an internal dashboard called Snapshot in late spring 2020. This was a valuable resource and reference tool for CDPH teams and leadership, as well as local health departments. It served as a centralized platform that compiled data from various sources, including test, case data, and hospitalization data, offering a comprehensive overview of county-specific information.
- Unlike the daily PowerPoint reports, which required intensive manual compilation, Snapshot would automatically update as the latest data became available. Because Snapshot was designed to be an internal-facing tool, the data needed less processing than the data reported on the public dashboards.
- Snapshot was most heavily used in late spring and early summer 2020 when the Blueprint for a Safer Economy (Blueprint) was implemented. LHJs were given access to Snapshot to inform their local policy decisions. The Blueprint provided a framework and established criteria that each county needed to meet before local businesses and economies could phase their reopening. Each week, counties were assigned to a COVID-19 risk-level tier based on updated data such as case rates and test positivity rates. CDPH released Blueprint Data Charts with tier assignments, case rates, testing positivity, and test rates. Given that the data indicated whether a county could reopen or not, CDPH's charts were heavily scrutinized. Sometimes LHJs disagreed with CDPH's determinations and



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

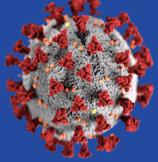
would request data adjudications to be placed in a different tier. For discussion of how the Blueprint was developed, see the Policy Development and Guidance chapter in this AAR. For further discussion of Blueprint data adjudication processes, see the County Monitoring and Local Coordination chapter in this AAR.

- Snapshot was deemed a major success and significant milestone in COVID-19 data automation. Since its creation, CDPH has made significant developments in data modeling, visualization, dashboards, and reporting with state, regional and local breakdowns, and as of summer 2022, Snapshot was still live and frequently used by epidemiology teams.

Data Workforce

Early Rotational Staffing Approach Abandoned

- Early in the pandemic response, CDPH's data and epidemiology teams established a normal emergency response structure, which is based on a rotational approach. In typical emergency responses, people work 12 hours a day, seven days per week, and rotate on and off after two weeks. However, it quickly became clear that this traditional emergency response staffing approach was not sustainable for data and reporting roles.
- For those teams working on data reporting and infrastructure, there was a heavy training component, which did not align with the amount of turnover associated with rotations. The teams found when new staff person had completed their training and gotten up to speed, their rotation would be over. Leadership soon realized that “we can't just do these kinds of complicated things with short rotations,” according to one SME. Especially for technical roles, institutional knowledge is critical, including understanding exactly how the data systems work. This takes longer than a short rotation to build this deep knowledge.
- In summer 2020, the data and technology teams abandoned the rotational staffing approach. CDPH focused on creating a COVID-19 response workforce through hiring and long-term redirections to work on the response indefinitely. This enabled staffing continuity and eliminated the upheavals associated with rotations. Ultimately, CDPH learned that rotational approaches might work well in some areas of pandemic response, but this staffing approach is not suited to the highly technical



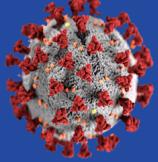
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

data and reporting workstreams. Durations of redirections and rotations should be reassessed for different roles in public health emergencies.

Data and Reporting Teams Were Understaffed and Overwhelmed

- During the first phase of the pandemic response in early spring 2020, CDPH's technology and data teams were understaffed and became quickly overwhelmed. As COVID-19 testing and case reporting increased rapidly, technical staff struggled to meet expanding data and reporting needs. There were no back-up staff, which meant that technical knowledge was often concentrated in a single person, who would quickly burn out. "You can't just have one person who knows the system, you need to have multiple people who know multiple aspects of that system," one SME noted.
- Additionally, the unprecedented daily reporting requirements and the scrutiny that data teams found themselves under contributed to low morale. Due to frustrations over work being constantly questioned, SMEs reported that many data and reporting staff who worked on the COVID-19 pandemic response did not volunteer during the mpox emergency activation.
- Looking back at these early challenges, SMEs noted several lessons learned. It would have been beneficial to have more robust and better-staffed data and reporting teams in place earlier on. These should include experts who understand technical nuances (such as ELR/HL7 reporting standards), as well as data analytics teams that can assess incoming data streams to quickly identify issues. While CDPH already employs individuals with these data and informatics skills, their numbers are very limited and are shared with different response efforts. SMEs also expressed concern that instead of investing in specialized skillsets such as data analysts, data managers, and informaticians, CDPH may fill these roles with epidemiologists and physicians. The pandemic demonstrated the prominence of data needed in an effective public health response, and the importance of hiring, retaining, and training staff who understand its complexities. The lack of this skillset is illustrated by the lack of an appropriate job classification for some of these skills within CDPH.
- Additionally, in future responses it will be critical to establish sustainable ongoing reassessments of reporting cadences proactively and strategically, and improved communications to data and epidemiology staff on how, why, and how frequently the data is used. Thoughtfully



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

considering the frequency, impact, and when to decrease the cadence for instance, would help mitigate burn-out in data and reporting staff, enable a more sustainable pace, and potentially produce more accurate reports.

- Despite the challenges associated with being under-resourced and understaffed, data and reporting teams created strong bonds across response teams and departments, which were defined by dedication, motivation, and camaraderie. The collaborations between data teams, epidemiology teams, and the CalREDIE program team was especially impactful. According to one SME, “a million little successes” were achieved throughout the response because of cooperation, collaboration, and ingenuity, including developing on-the-fly reports and ad-hoc problem-solving huddles. The experience of working with diverse colleagues in a heightened response environment has been “truly amazing,” another SME noted. The cross-functional and cross-disciplinary nature of response work has also helped break down siloes, which many agreed is a long-term benefit. Going forward, it will be important to sustain these relationships and momentum that were developed during the COVID-19 pandemic response.

Existing Staff Tapped to Help Recruit, Train, and Onboard New Hires

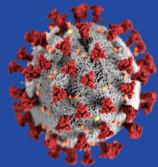
- Throughout the pandemic response, recruiting and onboarding new staff onto CDPH's data and reporting teams was challenging. “Onboarding has been really hit or miss,” one SME noted, explaining the significant investment of hands-on training that is necessary for new staff placed into technical roles. For the data and reporting teams, the responsibility to recruit, onboard, and train often fell to those who were engaged in operational response work. Ultimately, this diverted technical bandwidth away from their response work. Unfortunately, some new employees received minimal guidance since the experienced data analysts were diligently concentrating on operational response work. “It’s a challenge that the same people who are trying to turn out daily reports are expected to write job descriptions,” one SME commented. In future pandemics, it would be helpful to standardize hiring models across teams, as well as provide teams with administrative support to help with recruiting, onboarding, and training new staff.

Equity

This section describes equity considerations specific to this chapter.

As part of its unwavering commitment to reduce health inequities, CDPH launched a [Health Equity Dashboard](#) in November 2020 to track California's health equity measure and data by race and ethnicity, sexual orientation, and gender identity. This dashboard influenced the equity work that was promulgated by the Blueprint for a Safer Economy. The Blueprint, which tied counties' abilities to reopen their economy to meeting certain thresholds, inaugurated the State's focus on equity, which became increasingly emphasized over the course of the response. When it came to equity-related metrics, although the data was available internally to CDPH very early in the response, it was not released publicly until the equity dashboard was established. Leadership had to reach consensus on the data (some of which was sensitive and incomplete) knowing that it would be subjected to legislative and advocacy group scrutiny. Many conversations had to take place around how to present the data, what should be released, and how it should be made consistent with other dashboards. While the data was available and the dashboard was ready in Summer 2020, these leadership conversations and the associated approval processes took time. SMEs were frustrated by the belated release of case and death data by race, ethnicity, and location.

According to SMEs, when specific populations approached the State and requested COVID-19 data on their populations, the State delayed displaying it due to the instability of the data. For instance, California's Native Hawaiian and Pacific Islander community began asking for their data in Summer 2020. It was only after the community submitted a stakeholder letter to the California Legislature and the Governor's Office that CDPH was instructed to release their data. CDPH SMEs found this "highly problematic" and pointed out that if data by race and ethnicity is not shared early and often, local populations and LHJs have no way to know if a specific population is experiencing inequities. "We could have done more about that much earlier," one noted.

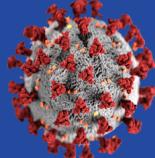


Data & Technology

This section describes data and technology specific to this chapter.

Test Data Volumes Destabilized CDPH's Existing Disease Surveillance System (CalREDIE)

- In mid-January 2020, the CalREDIE program team created a new disease condition for COVID-19 in the system. At this point, testing for COVID-19 was very limited and mainly controlled by the CDC. However, Statewide needs for COVID-19 testing quickly accelerated in early Spring 2020 as testing shifted from the CDC to individual states, and California began expanding testing capacity. As the State's disease surveillance system, CalREDIE collects and stores all data from labs, healthcare providers, and most local health departments for reportable disease conditions, including COVID-19. All 61 LHJs in California use CalREDIE in some capacity, and 58 LHJs use the system for surveillance of all notifiable communicable diseases.
- CalREDIE includes multiple technology components and applications, including a front-end electronic record reporting gateway. During ELR processing, the raw lab results are classified into several records (known as "incidents"), including a positive confirmed case, a probable case, an indeterminate case, or a negative case. Next, CalREDIE's systems and applications identify and quantify the records that meet the definition of a COVID-19 case—e.g., those with positive test results. All cases are incidents but not all incidents are cases.
- As cases started to rise and testing volumes grew in late spring 2020, the CalREDIE program team realized the system would soon be overwhelmed and began making modifications and expansions. The first component to be impacted was the front-end gateway, which receives raw lab data and initiates data processing on those messages. Before the pandemic, CalREDIE was processing an average of 10,000 test results per day. As testing ramped up, so did the system's daily averages. By mid-summer, the system was processing between 100,000 and 200,000 test results per day.
- As test data increased, CalREDIE began to slow down and was unable to reliably produce daily metrics. Overwhelmed by volume of data, early delays at the gateway created a ripple effect, impacting the system's ability to separate positive from negative test results. The counties then



CDPH COVID-19 After Action Report

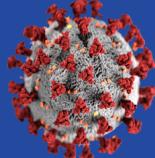
Chapter 19 – Data and Reporting

became impacted by this, since LHJs used CalREDIE to access their data. LHJs reports took four hours to generate instead of 30 minutes. Similarly, CDPH's epidemiologists who relied on the system faced significant delays: reporting teams would receive raw data and spend 24 hours generating a report but were increasingly unable to finish before the next day's data was available due to processing delays.

- Additionally, other CDPH programs that focused on different reportable diseases (e.g., tuberculosis) were also negatively impacted by CalREDIE's gradual failure. When other programs' access to their data grew inconsistent, CDPH moved to establish a separate, dedicated system for COVID-19 data.
- While CDPH initiated the procurement and implementation of a new technology solution to help alleviate the processing burden, it also took more immediate actions, including providing the data team first with "supercomputers," and then, when those failed, with dedicated SAS servers built by ITSD in a matter of days. The team collaborated heavily with ITSD for most of 2020, since "they had the power to provide us with space and functionality to pull data out," according to one SME. These short-term solutions helped the CalREDIE program team and system manage the increasing processing loads until the California COVID-19 Reporting System was implemented in October 2020.

Implementing a New Reporting System (CCRS) and Data Warehouse (Snowflake) Helped Improve Test Data Quality and Standardization

- Procured in late summer 2020, CCRS was comprised of two main components: an "upstream" gateway in front of CalREDIE to help standardize data and improve data quality, and a data warehousing component located "downstream" of CalREDIE to assist in data storage and reporting (also known as Snowflake and discussed further below).
- The new gateway placed in front of CalREDIE was supported by the CCRS solution vendor, Optum. As increasing numbers of labs began performing COVID-19 testing, the numbers of data quality issues data began to rise. Since the State has little control of the quality of lab data being sent in due to a lack of mandates, Optum helped interface with the labs to improve and standardize data quality prior to it being fed into CalREDIE. This effort included manually contacting every single lab to request corrections to poor quality data, as well as installing software to analyze incoming data quality and suggest corrections to missing fields.

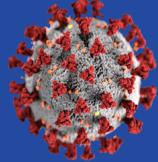


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

According to SMEs, the vendor contacted almost 500 labs who were sending duplicate or poor data.

- Ultimately, by scaling up CalREDIE and implementing CCRS to support and supplement it, the State was able to successfully meet the enormous test data volumes associated with COVID-19. During the Omicron surge, in late December 2021, the systems were able to process approximately 750,000 test results in a single day—a number that was “unheard of,” in the words of one SME. The CalREDIE program team, which at the start of the pandemic was comprised of approximately 20 people, grew to 50 by 2022. While CalREDIE’s failure was bolstered by the creation of CCRS, SMEs felt it should have been done sooner. If the public health systems had been regularly invested in, the State would have been better prepared for the pandemic. According to one leader, “information technology was a significant hindrance. We've been working hard to make it better, and it is better, but IT continues to be a rate-limiting factor in our ability to provide information.”
- Once CCRS had improved the quality of incoming test data, the State was able to focus on automating manual processes and making improvements. In Summer 2022, CDPH established parallel processing for ELR, established case registry processing in Snowflake, and began preparing for electronic case reporting (ECR), which will enable cases of reportable disease conditions to be transmitted from providers’ electronic health record systems.
- In 2022, CDPH also made several changes to CCRS, including carving out its data warehouse component and rebranding CCRS as the Surveillance and Public Health Information Reporting and Exchange (SaPHIRE). Whereas originally the CCRS data warehouse component, Snowflake, was managed by the vendor, in summer 2022 this was transitioned over to CDPH, who now manages and oversees the data warehouse in its own Snowflake instance. Snowflake houses and serves as the “source of truth” for all COVID-19 data, including test and case data, vaccination data, and therapeutics data. SMEs agreed that the implementation of Snowflake was a “100% game changer and huge success,” and that it drastically improved data quality, sharing, and management. These improvements have also been accompanied by a greater recognition of the importance and value of data systems to public health.



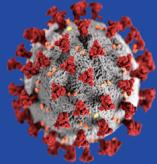
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

- Plans are currently underway to analyze alternatives and estimate costs for a new system to replace CalREDIE, a project known as the Future Disease Surveillance System (FDSS).

New Cloud-Based Contact Tracing System (CalCONNECT) Overwrote CalREDIE Data

- Following the announcement of California's Statewide contact tracing program, California Connected, in May 2020, one of CDPH's first tasks was to create a contact tracing system to enable the program. There was no existing software platform or system in place. The State conducted an expedited procurement process, selected an IT vendor, and the first iteration of the CalCONNECT system was launched in just 10 days on May 13, 2020. In summer 2020, the CalCONNECT team focused on integrating the new cloud-based system with the legacy CalREDIE system. This integration was technically problematic and the focus of an enormous amount of effort.
- In addition to the technical integration challenges, the interest, visibility, and funding associated with CalCONNECT caused problems with data integrity. The contact tracing program and system became a “driving force that steamrolled everything,” according to SMEs. The excitement and rapid timelines associated with the new CalCONNECT obscured the fact that CalREDIE remained the source of report for CDPH’s major COVID-19 metrics. According to one SME, “people didn’t understand that [CalCONNECT] was solely for contact tracing, and it wasn’t the source of report.” With many contractors and consultants working on CalCONNECT, not everyone was aware of the nuanced differences between the two systems. Consequently, CalCONNECT began overwriting certain data fields in CalREDIE, including death data. Death data fluctuated widely, and, until the error was discovered, caused CDPH data and reporting teams to spend time investigating the discrepancies. In the future, when new response systems are implemented and integrated, it will be important to carefully distinguish relationships between data streams to maintain data integrity.
- For further discussion of CalCONNECT and California Connected, see the Contact Tracing chapter in this AAR. For further discussion of CalCONNECT and CDPH’s other pandemic systems, see the Enterprise Technology chapter in this AAR.



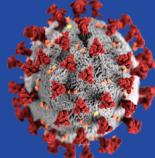
Communications

This section describes communications specific to this chapter.

Internal

Communications with Leadership

- Communication between data and reporting teams and CDPH leadership was problematic due to the lack of a formal Incident Command System (ICS) and emergency preparedness training. Without an ICS structure, the process for requesting information communication was informal and unstructured, and requests came in via email, Slack, and other chatting tools, phone calls, text messages, and other channels. There was little coordination around these requests, and data and reporting teams “felt like we had five different bosses all asking for the same thing,” according to one SME.” This lack of coordination caused teams to devote significant time responding to duplicate requests from leadership. It was difficult for teams to keep track of requests for information—and to discern whether all requests should be fulfilled. One leader noted that while staff and leaders are familiar with ICS, in emergency situations leaders often need information urgently and will go to multiple sources to obtain it quickly—unless there is a centralized location or source to which all requests for information can be directed.
- Over the course of the COVID-19 response, communications from leadership to data and reporting teams improved due to several factors. The automation of reports, enhancement of information systems, centralization of data and reporting processes, and establishment of innovative dashboards reduced the frequency of information requests. When Coronavirus Science Branch was reorganized into DISCO team, additional communication structures were established along with more regular check-in meetings.
- In future responses, many felt that CDPH should establish and follow ICS communication structures and channels, and discourage the use of informal, casual channels for decision-making or significant requests. This would require establishing a formal, internal process for information requests. “Decision-making shouldn’t happen inside casual conversations,” one SME noted. This lesson learned was reinforced by



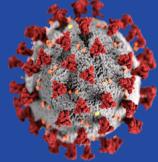
CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

many, who agreed that important communications should not happen “on the side” or “in texts.”

Communications across CDPH Response Teams and Programs

- Data and reporting communications across CDPH response teams and programs improved significantly over the course of the response involved different groups who worked with various datasets at different stages throughout the data lifecycle, from initial receipt to processing, dissemination, and analysis. In particular, the partnership between the CalREDIE program team (who managed incoming test result data), the data, reporting, and informatics teams (who cleaned, processed, and integrated data so it could be used), and the epidemiology team (who analyzed the data and used it to inform policy decisions) grew particularly strong. Though these teams each played different roles, they shared a common commitment to data integrity and to implementing solutions to improve California's data reporting.
- Communications with the Testing Task Force (TTF) were more challenging, due in part to the fact that it was stood up outside of DISCO and answered to a different authority. Since CDPH data and reporting teams owned the source testing data, they received frequent requests for data and tried to be responsive but were challenged by the fact that requests were often unclear, reflected changing priorities, and were poorly communicated. Data and reporting teams worked with the TTF data manager, who was a part-time redirected employee, and discovered that dashboards were being duplicated. “Data management with TTF was really hard,” one SME noted. It could have been improved with clearer guidance and communication from TTF leadership, better communication, and the addition of skilled workforce in data management and informatics to the TTF.
- In general, communications between teams began as ad-hoc and turned into standing meetings depending on needs. For instance, in mid-2020, CalREDIE team members met regularly with the contact tracing program as CalCONNECT was built and implemented. Most of the communication across response teams and programs consisted of “organic teamwork” (according to one leader) and grew out of the pre-existing relationships that CDPH staff, many of whom were redirected to the pandemic response, already had with each other. One leader



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

acknowledge that such ad-hoc coordination was harder for response team members who did not come from CDPH.

- Data and reporting management also had regularly scheduled virtual coordination calls with response leadership. These began as daily briefings, but over the course of the pandemic they decreased to five and then two times per week. These calls included other response teams and were helpful in helping management identify intersections between workstreams and stay up to date on policy and priorities. While the names and cadences of the calls changed, they remained throughout the years of the pandemic response.

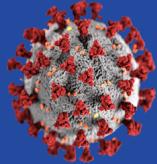
Communications between the CalREDIE Program and ITSD

- Prior to the pandemic, the CalREDIE program and those working on test and case data and reporting had limited interaction with ITSD, since there was only a single ITSD staff person assigned to them. The team was unsure of whom to approach for support and often felt “like we were on our own,” according to one SME. When, in summer 2020, data managers and analysts noticed signs of data system failures, the CalREDIE program found it difficult to elevate the issue and convey the criticality of the issue. As a result, the data failures were not immediately escalated and addressed as quickly as they could have been. The data failures led to a backlog of 250,000-300,000 test results, which led California to temporarily report inaccurate case numbers and case positivity rates.
- The lessons learned through this and other challenges resulted in an “extremely tight partnership between CalREDIE program and ITSD up and down the chain,” according to one leader. SMEs agreed that the strong, integrated relationship between the CalREDIE program and ITSD was a significant success and has benefited CDPH immensely.

External

Communications with the Public and the Media

- The CDPH Office of Communications would also communicate COVID-19 data to the public, respond to media requests, or make announcements on especially high-profile topics. In these situations, Office of Communications staff would work with data and reporting SMEs, including DPI, to better understand the request, comprehend the data, develop talking points, and assist with messaging. According to one SME, “they



CDPH COVID-19 After Action Report

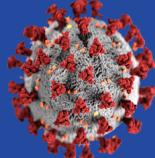
Chapter 19 – Data and Reporting

asked so many questions to really understand everything so they could take ownership of it, and take the burden off us." This data and reporting teams felt that when the Office of Communications was included in messaging and media responses, communications went much better.

- Some SMEs believed that from a public and media communications perspective, COVID-19 data was not always used as effectively as it could have been and was often reactive. The teams would receive requests for specific graphs or data to convey a specific talking point. Other times, requests were received to produce a graph similar to those presented by other States or jurisdictions. "It never seemed like the right order" to produce data graphs and charts that were ad hoc, one SME noted. Responding to these requests also disrupted work on longer-term projects.
- Additionally, many agreed CDPH was often overwhelmed by the volume of COVID-19 data and got too caught up in details and metrics. "We really need to have the data work for us, but we weren't using it to tell a good story," one SME noted. For instance, the plethora of available vaccine data obscured the most significant data point, which is "that vaccines save lives," another commented. The hyperfocus on real-time data and constantly reacting to requests contributed the struggle to communicate data strategically and meaningfully.
- In hindsight, teams acknowledged the missed opportunity to proactively use data to preempt potential requests and to develop consistent, straightforward public health messaging. As one leader reflected, "there was space for some consistent messaging, but we got distracted by the chaos and pressure." In the future, SMEs recommended developing a proactive strategy around data and reporting communications early in the response, which requires tight collaboration and establishing a "feedback loop" between the Office of Communications and data and reporting teams.

Communications with LHJs

- In the first 6 months of the response, CDPH lacked methods to effectively distribute processed case data to LHJs and this contributed to challenging communications. Counties could not replicate the State's COVID-19 metrics, and consequently CDPH SMEs needed to spend considerable time communicating how the case data was processed and calculated.
- During this time, LHJs downloaded their raw case data directly from CalREDIE's data distribution portal, but this dataset was not processed. On

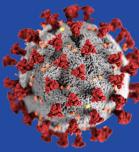


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

the State side, the CalREDIE team struggled to keep the system updated to meet changing reporting requirements. The team found it challenging to keep up with the pace and make sure that LHJs portal users had access to correct data in a timely manner.

- After CCRS and Snowflake were implemented, DPI granted access to LHJs so they could access their data both in raw and processed formats. However, the CalREDIE and DPI teams continued to support the LHJs through frequent communications. For over a year, the team provided user support seven days per week, in addition to holding at least two calls (MS Teams or Zoom meetings) per week including the CalREDIE Local Users call, CalCONNECT team calls, and other meetings. During these discussions, the CalREDIE and DPI teams would update LHJs on system and process changes and answer their questions. As of summer 2023, many of these calls still take place but with less frequency.
- Additionally, before auto-processing of positive test results was implemented in early December 2020, the CalREDIE program created teams to help LHJs who were falling behind on reviewing new records. These teams helped LHJs import and manage the data manually until CDPH received approval to turn on auto-processing. According to one SME, “with the team that we have, I think that the support we were able to provide LHJs users was amazing.”



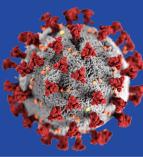
Workplan

This section is designed to be used as a workplan for future pandemics.

Definitions:

- **Phase:** The phase of the response in which the major tasks should be conducted (Planning; Initial start-up, Ongoing operations, or Close-out).
- **Major Tasks:** The tasks and activities that have to be conducted as part of the public health emergency response to a respiratory pandemic.
- **Success Criteria:** Criteria used to assess whether a task has been achieved successfully.
- **Considerations Based on COVID-19 Response:** Things to consider, including pitfalls, risks, and lessons learned, based on the COVID-19 response.
- **Finding ID:** The ID(s) from the related Finding/Corrective Action (where applicable).
- **Lead:** The lead person(s) responsible for task completion.

Phase	Major Tasks	Success Criteria	Considerations	Finding ID	Lead
Planning; Initial start- up; Ongoing operations	Identify and report on needed data streams, sources, and metrics strategically	• Data streams, sources, and metrics are useful, high-quality, and epidemiologically relevant.	• CDPH should be the source for all public health data. • Consider long-term and downstream impacts of decisions. • Take the time to think through all aspects of data	• Data and Reporting 3, 4, 5, 6, 23	

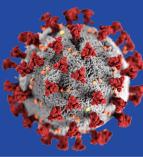


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Phase	Major Tasks	Success Criteria	Considerations	Finding ID	Lead
			<p>and reporting decisions.</p> <ul style="list-style-type: none">Once data and reporting patterns are established, recognize they can be difficult to change.Work with Cal OES on death reporting.		
Planning; Initial start-up; Ongoing Operations	Establish and maintain flexible, scalable data systems	<ul style="list-style-type: none">CDPH's public health data systems are flexible, scalable, and reliable.State leadership and policymakers can rely on public health systems for accurate, timely data.	<ul style="list-style-type: none">For new disease conditions, anticipate changing criteria, which necessitates system changes and affects previously submitted data.Consider data integrations and touchpoints between old and new systems.Sustain momentum on system improvements and implementations	<ul style="list-style-type: none">Data and Reporting 1, 7, 10, 19; Enterprise Tech 5	



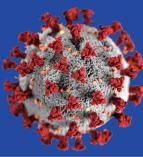


CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Phase	Major Tasks	Success Criteria	Considerations	Finding ID	Lead
Planning; Initial start-up; Ongoing operations	Establish and maintain technical data and reporting workforce	<ul style="list-style-type: none">• CDPH has adequate technical staff with the right skillsets to support all data and reporting.• Data and reporting teams are centralized and coordinated.	<p>already in progress.</p> <ul style="list-style-type: none">• Invest in a technical workforce with the right skillsets (informatician, data managers, data analysts, etc.).• Establish a centralized data and informatics Branch or Unit in the ICS structure that has insight into all data streams.• Create a Chief Data Officer or similar role.• Avoid a rotational staffing approach for technical staff.• Obtain administrative support to assist with recruiting, onboarding, and training of new staff.	<ul style="list-style-type: none">• Data and Reporting 2, 16, 17, 19, 20, 21	

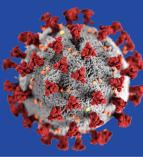




CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

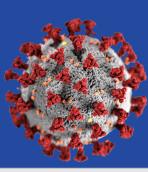
Phase	Major Tasks	Success Criteria	Considerations	Finding ID	Lead
Planning; Initial start-up; Ongoing operations	Establish and maintain consistent, clear communications	<ul style="list-style-type: none">• CDPH establishes a sustainable reporting cadence.• Data reporting messaging is consistent and proactive.	<ul style="list-style-type: none">• Establish and maintain ICS structure.• Maintain newly formed relationships between technical teams.• Ensure requests for information follow the appropriate channels so it gets to the right team.• Avoid decision-making in casual channels.• Set expectations around data reporting cadences, tolerance levels, and metrics.	<ul style="list-style-type: none">• Data and Reporting 8, 14, 15, 22	
Planning; Initial start-up; Ongoing operations	Manage expectations around the ultimate purpose of data reporting	<ul style="list-style-type: none">• Collected data is helpful to inform policy and decision-making.	<ul style="list-style-type: none">• Do not let the pursuit of high-quality data become an end in itself.• Release critical data and acknowledge that	<ul style="list-style-type: none">• Data and Reporting 11, 13	



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Phase	Major Tasks	Success Criteria	Considerations	Finding ID	Lead
			<p>it is imperfect and may change.</p> <ul style="list-style-type: none">• Question unreasonable reporting requests.		
Planning; Initial start-up; Ongoing operations	Establish data dashboards and portals to share information internally and externally	<ul style="list-style-type: none">• Data is shared timely and properly de-identified.• Stakeholders can easily access data and obtain answers to their questions.	<ul style="list-style-type: none">• Do not publish data that has not been properly suppressed or de-identified.• For public-facing dashboards, establish data dictionaries.• Internal data dashboards may present rawer, less curated data.	<ul style="list-style-type: none">• Data and Reporting 9, 15	
Planning; Initial start-up; Ongoing operations	Monitor and maintain data quality	<ul style="list-style-type: none">• Reported data is high quality.• Manual data cleaning is minimized.	<ul style="list-style-type: none">• Data received from third parties (including LHJs, providers, and labs) will vary in quality and completeness.• Manual “clean teams” may be necessary to help improve data quality.	<ul style="list-style-type: none">• Data and Reporting 11, 18	



CDPH COVID-19 After Action Report

Chapter 19 – Data and Reporting

Phase	Major Tasks	Success Criteria	Considerations	Finding ID	Lead
			<ul style="list-style-type: none">• Consider provider/lab outreach to help improve incoming data quality.		

DRAFT