



# 2022 U.S. DIGITAL SERVICE/ CDC CASE SURVEILLANCE DISCOVERY SPRINT

Findings and Recommendations

August 12, 2022

## Introduction

As part of CDC's Data Modernization Initiative (DMI), a team of 5 U.S. Digital Service (USDS) technologists and 2 CDC epidemiologists examined case surveillance through a 12-week discovery sprint from May to July 2022. This sprint sought to

1. understand the current state of case-based reporting and notification;
2. identify potential approaches to optimize policy, processes, and technologies; and
3. make actionable recommendations to advance and modernize case surveillance.

The sprint focused on case reporting and notification through the National Notifiable Diseases Surveillance System (NNDSS). To understand the current state of case surveillance, sprint research consisted of three phases:

1. desk research to understand case surveillance, prior reports and evaluations, and surveillance strategy for a sample of specific conditions;
2. interviews with CDC program representatives, surveillance technology subject matter experts, and thought leaders throughout CDC, Council of State and Territorial Epidemiologists, and National Association of County and City Health Officials; and
3. in-depth research with public health staff at 16 state and local health departments and overview conversations with 3 additional jurisdictions.

Surveillance data, technology, and tools have evolved significantly, as have the expectations on data quality, timeliness, and representativeness for both daily surveillance and emergency response. Outlined below, the sprint's findings have identified several opportunities for improvement and recommendations for developing the next generation of case surveillance across public health partners.

## Findings

The sprint found a complicated and resource-intensive system with numerous, sometimes conflicting, goals that state and local health departments find difficult to meet with the current workforce, systems, and processes. The resources spent on counting, cleaning, and collating case data are significant but do not always meet the needs of the system's numerous users, such as jurisdictions, CDC programs, or emergency responses.

### Opportunities for improvement

1. **Center on the needs of state, tribal, local, and territorial (STLT) health departments and communities.** National, state, territorial, local, and tribal public health organizations may have different objectives for disease surveillance, public health action, and case reporting. Many systems and processes in use today are misaligned with the needs on the ground at local health departments, where data are collected and urgent public health action is taken. The goals of case reporting should be defined, considering the various end users of data and factoring in the needs of and implications to STLTs. This should be better understood and reflected at the national level when designing systems, funding data collection, or evaluating burden.

2. **Account for the realities of the STLT workforce.** The STLT workforce required to update and maintain the current case surveillance system is difficult to hire in many jurisdictions. Problems range from pandemic-related burnout, attrition of people who know how to operate the current system, highly specialized informatics or IT needs, and systematic barriers to hiring. There is an opportunity to rethink and simplify how the jurisdictions are asked to provide case surveillance data to CDC so that they support the goals of both CDC and STLTs.
3. **Streamline the data pipelines.** Both at the jurisdictional level and at CDC programs, time and resources are spent solving the same problems when ingesting, validating, cleaning, parsing, and merging data and when extracting data for reporting or analysis. Some of these solutions are innovative and automated; some are labor-intensive and manual. In addition, this complexity slows down the transmission of data and erodes trust in the accuracy of the data. Most users receive data that crossed many system boundaries and underwent repeated transformations.
4. **Understand the difficulties involved in data collection and design systems, processes, and methods to address that.** Beyond core case data needed to identify cases, clusters, outbreaks, or trends, supplemental disease-specific data are burdensome and impractical for STLTs to collect, collected inconsistently leading to low completeness rates, and not always perceived as useful. A pandemic-ready system should accept that data are inherently messy and design systems, processes, and scientific methods accordingly; focus resources on leveraging data that exist or are easy to collect; and collect additional information judiciously and with a clearly communicated purpose.

## Recommendations

The sprint recommendations are focused on addressing three goals:

1. Support jurisdictions in collecting and sending case data in a timely manner, at a cost commensurate with the utility of these data.
2. Provide jurisdictions with a single transmission mechanism for all the data that leave their electronic disease surveillance systems, including disease-specific supplemental data and emergency response or outbreak data.
3. Provide value to national public health, leveraging unique opportunities that national data aggregation offers.

### Goal 1

***Support jurisdictions in collecting and sending case data in a timely manner, at a cost commensurate with the utility of these data.***

When it comes to sending NNDSS data to CDC, the goal of a jurisdiction is to transmit it accurately and without undue burden. Burden can be incurred throughout the case report

lifecycle, and CDC has the ability to worsen or relieve this burden at many of the lifecycle milestones.

#### Vision

- There is clear, accountable leadership at CDC in creating and stewarding national case reporting data standards.
- There is alignment across CDC standards, national healthcare standards like United States Core Data for Interoperability (USCDI), and efforts like electronic case reporting (eCR).
- Jurisdictions send standardized core data for all national notifiable conditions using the national data standard and a data format and transmission method that is most appropriate to their system.
- CDC, through efforts like DMI and North Star Architecture, uses its national scale to streamline and provide agile, modular, and adaptable tools to help electronic data processing, such as electronic laboratory reporting (ELR) or ECR, at the jurisdiction level.

#### Specific recommendations

1. Designate responsibility, authority, and leadership for the CDC Division of Health Informatics and Surveillance (DHIS) to create and steward national case reporting data standards.
2. Turn the current Generic v2 Message Mapping Guide (MMG) into a format-agnostic data standard. Onboard all jurisdictions to enable submission of a generic, core data feed that can be used for national disease surveillance. Expand the range of available data formats beyond HL7v2, such as comma-separated values (CSV) or Fast Health Interoperability Resources (FHIR).
3. Pause and rethink the current way of reporting supplemental case data ("condition-specific MMGs").
4. Leverage NEDSS Base System (NBS) modernization and CDC "building blocks" work that is developing functionality that will be used across diseases and conditions. Pilot innovative data ingestion and data processing tools with a few jurisdictions.

#### Goal 2

***Provide jurisdictions with a single transmission mechanism for all the data that leave their electronic disease surveillance systems, including disease-specific supplemental data and emergency response or outbreak data.***

At the jurisdiction level, most of the case-related data are stored and managed in a single integrated disease surveillance system, sometimes supplemented by a small number of disease-specific systems (such as for HIV, STDs, etc.). Currently, these integrated disease surveillance systems send the same data to multiple locations at CDC in multiple formats. Also, situations like outbreaks, emergencies, or enhanced surveillance funding often create new ways of

collecting, storing, and transporting data. Jurisdictions have expressed a wish to “send data once, to one place.”

### Vision

CDC should build a single, efficient, flexible and scalable data connection between jurisdictions and CDC for both routine public health and emergencies that supports structured, semi-structured, unstructured, and aggregated data. Case data, and any additional data that can be connected to cases such as full labs or questionnaires, should come to CDC through a single front door, and a well-resourced organization should maintain this connection.

In routine public health surveillance, jurisdictions send probable and confirmed cases of national notifiable diseases and any other case data they have agreed to or been funded to send. The same system should be used for the core data necessary to conduct national surveillance and for any supplemental data that may be collected in situations that warrant it.

In cases of urgent public health importance, a jurisdiction could choose to send suspected cases immediately, regardless of whether the condition is currently national notifiable, followed by subsequent updates with additional information. The system should support transmitting both pre-defined and emergent structured, semi-structured, and unstructured data attached to cases. When needed, the system should also allow jurisdictions to send aggregated case counts. The goal is to both provide timely counts to quantify the spread, trends, and burden and access to in-depth case investigation information to further characterize the disease or outbreak.

Over time, this system could be expanded to support other types of surveillance such as sentinel or syndromic.

### Specific recommendations

1. Develop a single entry point at CDC to which all jurisdictions send their case data.
2. Provide assistance to support jurisdictions in setting up these data flows.
3. Execute a proof of concept exercise through the NBS modernization project with a few diverse jurisdictions to prove that new conditions and new data elements could flow swiftly, using the same mechanism and CDC entry point, in case of an emergency.
4. Define the product strategy, software architecture, and organizational structure to implement this vision. Staff this workflow with expertise in human-centered design, product management, and software architecture.

### GOAL 3

***Provide value to national public health by leveraging unique opportunities that national data aggregation offers.***

Jurisdictions provide data to CDC on a voluntary basis based on a shared agreement to track high-importance diseases and conditions because of the positive impact on the nation’s health. National data collection also provides additional opportunities to create value thanks to its

scale, such as data aggregation, comparison, sharing, or enrichment, enabling jurisdictions to use NNDSS in a way that directly benefits their local communities.

Above all, as the leading federal public health authority, CDC has a responsibility to provide timely national situational awareness and to disseminate and share data used by many to inform resource allocation and decision-making.

### Vision

There should be an accessible, transparent, and secure data processing pipeline at CDC that makes timely data available to jurisdictions, CDC stakeholders, and systems in many forms.

All parties should have access to the raw format the data came in, cleaned and transformed data that are ready for analysis, or aggregated data. In this system, data would go through multiple data quality tiers, and both jurisdictions and CDC program users would have access to all tiers. The case data should be automatically available to any other analytical tools that can use them such as ESSENCE, SEDRIC, or DCIPHER.

Easy access to their own data that have been cleaned, aggregated, analyzed, and visualized would save jurisdictions resources. Jurisdictions at all levels could use tools to compare and understand public health threats in different locations around the country and support development of public health policy and legislation. Data that are collated and aggregated at a national level also can be enriched with other national data sets, such as census tract data or weather data.

### Specific recommendations

1. Define the product strategy and implementation plan with subject matter experts from DHIS, CDC Office of the Chief Information Officer, and USDS.
2. Pilot a few innovative data products that can be accessed in a shared environment for use by jurisdictions and CDC programs.
3. Set expectations around data sharing:
  - a. Clearly define data ownership and stewardship roles.
  - b. Develop a better understanding of data users, both inside and outside of CDC.
  - c. Describe what it looks like for data to be shared with the public in a timely, easily analyzed, and easily found manner and commit to it.

## Conclusion

These recommendations are not a blueprint but, rather, a foundation for developing a strategy and an execution plan, guided by principles of human-centered design and agile development, for developing next-generation case surveillance.

The over-arching recommendation of this sprint is to define clear leadership roles and responsibilities for case surveillance at CDC, with a clear mandate to serve specific, well-defined users inside and outside of the agency and empowered to deploy solutions to those users' problems iteratively.