AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Perspective

Improving efficiency of COVID-19 aggregate case and death surveillance data transmission for jurisdictions: current and future role of application programming interfaces (APIs)

Diba Khan[1], Meeyoung Park[1], Samuel Lerma[1], Stephen Soroka[1], Denise Gaughan[1], Lyndsay Bottichio[1], Monika Bray[1], Mary Fukushima[1], Brooke Bregman[2], Caleb Wiedeman[3], William Duck[1], Deborah Dee[1], Adi Gundlapalli[1], and Amitabh B. Suthar[1]

[1]Coronavirus Disease Response, Centers for Disease Control and Prevention, Atlanta, Georgia, USA, [2]California Department of Public Health, Sacramento, California, USA, and [3]Tennessee Department of Public Health, Nashville, Tennessee, USA

Corresponding Author: Coronavirus Disease Response, Centers for Disease Control and Prevention, Atlanta, GA, USA; ild1@cdc.gov

ABSTRACT

During the coronavirus disease-2019 (COVID-19) pandemic, the Centers for Disease Control and Prevention (CDC) supplemented traditional COVID-19 case and death reporting with COVID-19 aggregate case and death surveillance (ACS) to track daily cumulative numbers. Later, as public health jurisdictions (PHJs) revised the historical COVID-19 case and death data due to data reconciliation and updates, CDC devised a manual process to update these records in the ACS dataset for improving the accuracy of COVID-19 case and death data. Automatic data transfer via an application programming interface (API), an intermediary that enables software applications to communicate, reduces the time and effort in transferring data from PHJs to CDC. However, APIs must meet specific content requirements for use by CDC. As of March 2022, CDC has integrated APIs from 3 jurisdictions for COVID-19 ACS. Expanded use of APIs may provide efficiencies for COVID-19 and other emergency response planning efforts as evidenced by this proof-of-concept. In this article, we share the utility of APIs in COVID-19 ACS.

Key words: COVID-19, surveillance, CDC, application programming interfaces

Timely disease surveillance has been essential for decision-making to inform timely and appropriate actions during the coronavirus disease-2019 (COVID-19) public health response. The Centers for Disease Control and Prevention (CDC) has established a system for tracking COVID-19 aggregate case and death data, known as aggregate case and death surveillance (ACS). ACS publishes official daily counts of COVID-19 cases and deaths and enables timelier detection of trends for public health decision-making compared to tra-ditional case reporting. CDC works with state, territorial, District of Columbia (DC) and New York City (NYC) public health jurisdictions (PHJs) in a multistep, voluntary, process to collect data and confirm daily case, and death numbers for COVID-19.[1] ACS was initially designed to monitor the cumulative and daily number of COVID-19 cases and deaths recorded by the submission date—the date CDC collects information from a PHJs official website or counts reported by PHJs to CDC. COVID-19 case and death data

for ACS originate from PHJ-mandated reporters such as laboratories, health care providers, long term care and correctional facilities, vital records, or through contact tracing activities, etc. PHJs receive COVID-19 case and death data through their respective surveillance systems, which are then prepared and managed by PHJs for access by the public through their official websites. CDC collects daily jurisdictional COVID-19 cumulative case and death counts via an automated collection of web-based publicly available information, such as from state dashboards, county websites, media articles, and Epi Info™ surveys submitted on a voluntary basis by PHJs. CDC manages the flow of COVID-19 case and death data via an electronic surveillance system, the Data Collation and Integration for Public Health Event Response (DCIPHER), into a common operating platform (HHS Protect) for storing, aggregating, and sharing public health data.[2,3] Some PHJs have provided application programming interfaces (APIs) to published COVID-19 case and death data, which are available to the public. CDC has used APIs that meet data requirements for ACS to streamline data collection. APIs are a common technology used across different sectors to allow efficient machine-to-machine communication using established data standards.[4] APIs enable different systems, applications, and platforms to connect, synchronize and share data with one another for accurate, and efficient data transfer. Here, we share the utility of APIs in COVID-19 ACS.

Since Fall 2020, aggregate COVID-19 case and death data collection at CDC occurs through 2 parallel processes: (1) daily updates to collect new COVID-19 case and death data and (2) historical updates to integrate and reconcile retroactively reported COVID-19 cases and deaths to the correct date in the time series. The historical corrections to COVID-19 aggregate case and death data have increased in order to reconcile backlogged cases to the appropriate report date, and also due to data cleaning or retroactively applied case definitions. CDC collects historical COVID-19 case and death timeseries data from PHJs through a manual or automated process. In the manual process, the PHJs email a templated MS Excel file containing historical COVID-19 case and death timeseries data to a state coordinator at CDC. Certain PHJs have also enabled direct uploads of COVID-19 case and death time series data using comma-separated values (CSV)/Microsoft (MS) Excel files, which CDC can access and import into HHS Protect, provided the cases and deaths are summarized by daily report date and include confirmed and probable cases and deaths as separate values (if reported as such on their website or directly to CDC). CDC ingests new COVID-19 case and death data through to production, then validates and pushes analyzed data to platforms, such as the CDC COVID Data Tracker and to other data surveillance products.[5,6] As of March 2022, 28 PHJs have updated historical data at least once; 10 PHJs regularly submit the historical data to correct aggregate-level COVID-19 cases and deaths.

To update historical COVID-19 case and death time series timeseries using an automated process, PHJs can submit the data directly into CDC's data repository via APIs. An API consists of a defined set of instructions by which the software applications communicate with each other and enables data transfer between different systems through an established and well-documented interface. APIs facilitate the data transfer of daily and/or cumulative counts of COVID-19 cases and deaths from jurisdictional databases. CDC provides specifications to PHJs for API endpoints and standardized formats in which to provide COVID-19 case and death data. In the automated process using APIs, CDC can schedule COVID-19 case and death data pulls directly from PHJs into HHS Protect without the need for CDC or PHJ staff to manually perform initial validation and data upload (Figure 1). The COVID-19 case and death counts are further cleaned and prepared using data pipelines in HHS Protect. CDC conducts quality control validation checks before the COVID-19 case and death data are pushed to production platforms,
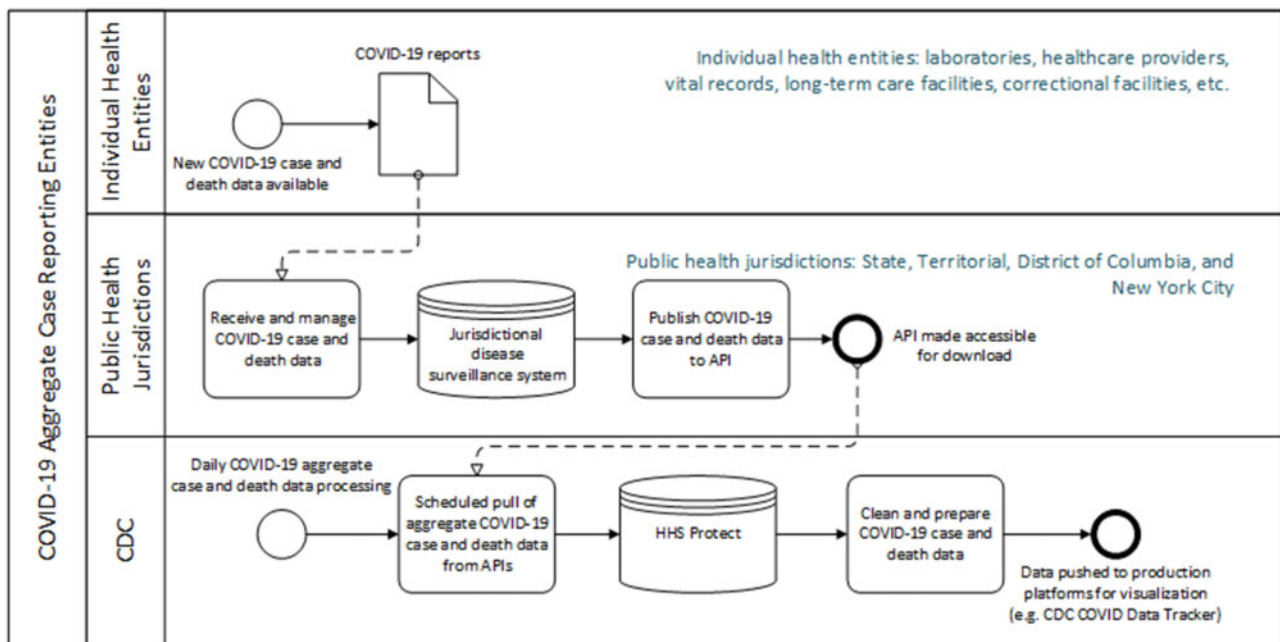

Figure 1. Data ingestion process from public health jurisdictions (ie, California Department of Public Health, Florida Department of Public Health and Tennessee Department of Public Heath) via application programming interfaces.

such as the CDC COVID Data Tracker and to other data surveillance products.[5,6] As of March 2022, CDC has transitioned to automatically pull current and historical COVID-19 case and death data from 3 PHJs, California, Florida, and Tennessee, using their APIs.

Automatic ingestion of current and historical time series for aggregate COVID-19 cases and deaths using APIs has had several benefits. Firstly, APIs improved efficiency by reducing the overall processing time from approximately 30 min to less than 5 min due to fewer steps requiring human intervention. Use of APIs alleviated the burden on PHJ and CDC staff by eliminating manual file transfers and data uploads and minimizing manual data validation steps. Secondly, APIs improved data quality by reducing opportunities for human error by using automated processes and reporting. Per testimony from Tennessee Department of Public Health personnel, "the main appeal of going to an API as the middleman was the ability to incorporate 100% automation on our side." Thirdly, APIs are highly adaptable for data transfer and are implementable across a range of surveillance systems, and therefore can save on PHJ's developmental costs. Finally, APIs help ensure timely and accurate situational awareness of PHJs COVID-19 case and death data while allowing PHJs to have full control of the data being viewed. Ownership and stewardship of COVID-19-related data remains with PHJs, allowing PHJs to update the data as needed without added intervention from CDC. Hence, an API-based approach to pandemic surveillance provides both more accurate reporting and rapid incorporation of changes to COVID-19 case and death data when compared with the manual approach as discussed above. Collectively, these benefits enable CDC to more frequently refresh COVID-19 case and death time series data than would be possible using only manual updates, and thereby more closely align CDC's data to data presented on PHJ's websites on a real-time basis. APIs may also have a potential benefit in obtaining new surveillance data from jurisdictions on a daily basis. These API benefits may contribute to future pandemic and other emergency response planning efforts by improving collaboration with PHJ's and other data modernization initiatives.[7]

PHJs may require technical support, resources, and time to build and publish their APIs, which requires a fixed initial investment in exchange for reduced maintenance once API connections are set up and validated. PHJs may not have implemented APIs for data reporting due to the lack of resources or expertise in setting up an API; establishing data security and privacy measures will be required prior to operationalizing APIs. Standardized specifications with defined queries and variables for APIs could facilitate data transfer to CDC not only during public health emergencies but also for routine surveillance. Our experience in using APIs for ACS may aid the development of a generic aggregate surveillance system for future public health emergencies. CDC guidance on technical considerations include shared definitions and vocabularies; aligned processes such as frequency of reporting, availability of PHJ resources to setup an API, and to regularly refresh the dataset, which may be helpful to expand use in more jurisdictions.

AUTHOR CONTRIBUTIONS

DK, MP, SL, ABS, BB, CW, DG, DD, WD, SS, LB, MF, MB, and AG contributed to the initial conception and design of the article. DK, MP, SL, BB, CW, SS, LB, and ABS provided detailed content related to drafting of the article. DK, MP, SL, ABS, BB, CW, DG, DD, WD, SS, LB, MF, MB, and AG helped in revising it critically for important intellectual content. All authors reviewed and approved the submitted manuscript and have agreed to be accountable for its contents.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

There are no new data associated with this article.

DISCLAIMER

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. Council of State and Territorial Epidemiologists. Update to the Standardized Surveillance Case Definition and National Notification for 2019 Novel Coronavirus Disease (COVID-19). 2021. https://cdn.ymaws.com/www.cste.org/resource/resmgr/21-ID-01_COVID-19_updated_Au.pdf Accessed November 3, 2021.
2. Palantir Media. CDC Renews Partnership with Palantir for Disease Monitoring and Outbreak Response. 2021. https://www.palantir.com/newsroom/press-releases/cdc-renews-partnership-with-palantir-for-disease-monitoring-and-outbreak/ Accessed April 6, 2022.
3. HHS Protect Public Data Hub. 2021. https://protect-public.hhs.gov/ Accessed April 6, 2022.
4. Centers for Disease Control and Prevention. APIs. 2021. https://open.cdc.gov/apis.html Accessed October 13, 2021.
5. Centers for Disease Control and Prevention. COVID Data Tracker. 2021. https://covid.cdc.gov/covid-data-tracker/#datatracker-home Accessed September 26, 2021.
6. Centers for Disease Control and Prevention. COVID-19 Case Surveillance Public Use Data. 2021. https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf Accessed October 29, 2021.
7. Centers for Disease Control and Prevention. Data Modernization Initiative. 2021. https://www.cdc.gov/surveillance/surveillance-data-strategies/data-IT-transformation.html Accessed October 13, 2021.