The Data Coalition Initiative advocates for policy reforms that enable the government to produce accessible, high-quality and useful data. As we continue to grapple with and recover from the COVID-19 pandemic, the value of a robust data infrastructure cannot be understated, especially when it comes to understanding social determinants of health (SDOH), which are the social and environmental factors that shape health outcomes. Understanding how populations that are vulnerable to health crises like the COVID-19 pandemic are impacted by SDOH, may help public health officials, epidemiologists, and policy decision makers study and make well informed intervention recommendations. Yet, these efforts are limited by the inability to effectively link data across multiple disciplines and government silos. These data constraints include a lack of structured or standardized data, legislative barriers, and outdated IT systems. The ability to confront the current pandemic and get ahead of future ones depends on the United States’ ability to modernize health related data infrastructures.
Current Health Data are Insufficient to Address the Covid-19 Pandemic

It is now widely accepted that social determinants of health (SDOH) have a major impact on short and long term health outcomes, care quality, and medical costs. Understanding and analyzing data on SDOH provides policymakers with the evidence needed to make public health interventions effective and efficient in improving community health across all communities.

The COVID-19 pandemic is having a disproportionate impact on vulnerable populations, due in significant part to factors that are not considered in a traditional health care setting. For example, the ability to work remotely, reliance on public transportation, and ability to isolate should one become sick all impact a person’s likelihood to contract and recover from COVID-19. There are still significant structural barriers to getting appropriate data of sufficient quality to researchers and other stakeholders to help design effective policy interventions that take into account the complex nature of SDOH.

With improvements to data standards, improved data sharing mechanisms, and investment in modernizing systems, the United States can create a data infrastructure to improve policies, offer deeper, more timely analytical insights, and greater participation.

Lack of standardized, structured data inhibits understanding health.

With the growing interest in measuring SDOH and their effect, stakeholders are working to produce additional measures of different health risk factors and indicators. However, these efforts are non-standard. Without any guidance from the federal government, jurisdictions decide on their own how to compile and present information, creating hundreds of uncoordinated systems, resulting in inconsistent and incomparable data. This impacts the quality and availability of national level data, which, in turn, keeps policy makers and public health officials from prioritizing and executing effective responses.

This inability to make quality comparisons between jurisdictions also creates challenges for researchers looking to create insights into what effective and efficient interventions may be like in other scenarios.

For example, it is not possible to compile precise national summaries of COVID-19 impacts for any race or ethnicity category, or to make comparisons across state lines. Alaska reports nine race and four ethnicity categories, while West Virginia reports four race categories and no ethnicity categories. This lack of standardized, comparable data makes it difficult to assess regional and national public health challenges and develop reasonable policy responses.

Furthermore, a lack of standards can mean that some crucial data may be missing from reporting data. Again, taking race and ethnicity as an example, a state may have data that identifies individuals as belonging to racial and ethnic categories, however, it does not mean they are using those definitions consistently across all reports. Race or ethnicity may be captured on official reports of hospitalizations and deaths, but not vaccinations.

Data Standardization:

Enables data sharing and access.

Creating a capacity to securely share information while protecting confidentiality and deploying other privacy safeguards offers tremendous potential for developing new insights and knowledge to support statistical analysis and summary level information relevant for evidence-based policymaking and practice. For example, the National Center for Health Statistics (NCHS) linked their National Health Interview Survey and their National Health and Nutrition Examination Survey with administrative data from the Department of Housing and Urban Development’s largest housing assistance programs, including the Housing Choice Voucher program. This linkage program has been cited by several researchers, including those examining the blood lead levels of children in public housing.

Allows for the creation of structured, machine readable data.

Structured data can help improve timeliness and reduce reporting errors, as well as allow for the rapid development and use of improved analytical techniques. Structured health data allows opportunities to link these data to other types of standardized data, like census, education, labor, and transportation data, creating better insights into how health interacts with other important factors in a person’s life.

Permits open data to be published.

Publishing public health information as open data or free, publicly available information allows for better and wider access among community stakeholders, policy makers, and researchers. Open data has the potential benefits of increasing opportunities for scientific collaborations, enriching research and analytical capacity, and improving equity of access to data about communities who may not otherwise have access to restricted data files (including compliance with the American Disabilities Act).

¹ The social determinants of health (SDOH) include social and political factors such as socioeconomic status, race, ethnicity, the physical environment in which people live, and a major impact on health outcomes, care quality and medical costs (https://health.gov/healthypeople/objectives-and-data/social-determinants-health).
Legislative Solution to Identified Data Challenges

The Health STATISTICS Act (H.R. 831) builds on the unanimous recommendations of the U.S. Commission on Evidence-Based Policymaking and subsequent bipartisan legislation, the Foundations for Evidence-Based Policymaking Act of 2018. The proposed bill would modernize certain evidence-building activities for public health by improving access to high-quality data in the most secure manner possible. Specifically, the Health STATISTICS Act will:

- Require HHS to share timely, accurate, and actionable public health information collected by the Department of Health and Human Services with the Centers for Disease Control and Prevention (CDC), other public health agencies, and the public, to the extent possible while preserving individual privacy.
- Require the CDC, in collaboration with relevant agencies, to designate data and technology standards in order to streamline data reporting and collection requirements for public health reporting entities and develop common standards for the highest priority data elements to ensure that public health officials have access to comprehensive and data that can inform decision-making in near-real-time.
- Create a grant program for State, local, Tribal, or territorial public health departments for the expansion and modernization of public health data systems, to assist public health departments.
- Direct the National Center for Health Statistics to establish and implement a data systems linkage program of appropriate federal statistical data sets and establish a Health Data Evidence-Building Fund for developing data linkage and IT infrastructure services for public health researchers.
- Require the National Center for Health Statistics to conduct a demonstration program on SDOH and data linkages across federal agencies.

Making Progress

During the Pandemic, researchers noted the difficulty in dealing with rapidly changing data during a global public health emergency. The effects of policy interventions to slow or stop the spread of COVID-19 were difficult to measure because of the rapidly changing data landscape, the lack of data automation, the lack of open source data, and the varied sources of data including official government reporting, social media accounts, cell phone location data, media reports, AI diagnostic data, and open source datasets¹. Programs that seek to address SDOH lack the data necessary to make practical, evidence-based decisions. Improving data infrastructure can help support programs that rely on such information to design effective and efficient policy interventions. Data infrastructure can be improved by creating opportunities for data linkages to support research and evaluation, supporting responsible open data, and supporting state and local entities to innovate.

Adopting standards and creating evidence is already a priority of the Centers for Disease Control and Prevention (CDC) and other agencies. By urging the adoption of standards, Congress can further support data for public health. There are several successful approaches to linking data while ensuring privacy. Other examples include:

- America’s DataHub Consortium, a pilot program at the National Science Foundation, is exploring how to link data, foster innovative data structure, ensure privacy, and analyze data in new ways.
- The National Center for Health Statistics’ data linkages to the National Death Index, CMS Medicare and Medicaid data, US Renal Data System, Social Security Administration, and the Department of Housing and Urban Development.
- The Health and Retirement Study collects data related to states and their COVID-19 policies and outcomes, particularly as they relate to vulnerable populations.

The Role of Technology in Addressing these Challenges

Many public health labs and state public health departments struggled with legacy IT systems prior to the pandemic. So, when the demand for data increased, data collection and reporting were hampered by paper reporting. Electronic data reporting has the potential to reduce the burden on front line workers and laboratories as well as make sure data is reported faster and is less prone to errors. Often these entities rely on outdated reporting mechanisms because there is a lack of funding to support improvements and concerns over regulatory compliance.

Additionally, new privacy preserving technologies such as multiparty computation and homomorphic encryption hold the promise of being able to share sensitive health data in a way that preserves privacy and confidentiality.

² [https://experience.arcgis.com/experience/a29e56ff1b44cba3c2d1134354074/](https://experience.arcgis.com/experience/a29e56ff1b44cba3c2d1134354074/)
³ [https://dhhr.wv.gov/COVID-19/Pages/default.aspx](https://dhhr.wv.gov/COVID-19/Pages/default.aspx)
⁴ [https://www.nature.com/articles/s42254-020-0178-4](https://www.nature.com/articles/s42254-020-0178-4)