

# RECOMMENDATIONS

JULY 2020 • GENOMIC EPIDEMIOLOGY DATA INFRASTRUCTURE NEEDS FOR SARS-COV-2: MODERNIZING PANDEMIC RESPONSE STRATEGIES

## RECOMMENDATION 1

The U.S. Department of Health and Human Services should ensure the generation of representative, high-quality full genome sequences of SARS-CoV-2 across the United States, and in the future, from emerging epidemic or pandemic pathogens, in order that these data can be used to meet key needs for genomic surveillance

- Pathogen samples must be obtained from individuals who represent broad diversity of factors such as race and ethnicity, gender, age, geography, and other demographic features such as housing type, clinical manifestations and outcomes, and transmissibility.
- Capacity for genomic sequencing should be developed and supported at many geographically distributed sites performing testing, including public health laboratories, academic and medical centers.
- Representative SARS-CoV-2 clinical samples from across the United States should be collected and sequenced on an ongoing basis to provide baseline data and facilitate near real-time transmission tracking.

Genome sequences should be shared openly on publicly accessible databases, such as the National Center for Biotechnology Information linked to the Global Initiative on Sharing All Influenza Data.

## RECOMMENDATION 2

The U.S. Department of Health and Human Services should develop and invest in a national data infrastructure system that constructively builds upon existing programmatic infrastructure with the ability to accurately, efficiently, and safely link genomic data, clinical data, epidemiological data, and other relevant data across multiple sources critical to a public health response such as the current SARS-CoV-2 outbreak. Such a system should:

- Allow for the linkage of genomic data, clinical data, epidemiological data, and other relevant data in a way that is not overly burdensome to laboratories that collect data regularly.
- Create and foster safe data sharing practices to ensure that individuals' personal identifying information remains unexposed when data are being used and shared across the system.
- Be grounded in the pursuit of standardization, interoperability, flexibility, and the practical linkage of data, including consideration of a potential national patient identifier.
- Consider not only the data required to create such a system, but also investment in mechanisms supporting the collection and analysis of such data, including promoting formal education in "data wrangling" at the intersection of data science and infectious disease epidemiology.
- Conduct regular annual reviews—including scenario-based simulations—to identify capacity gaps, promote process improvement (based on existing U.S. infrastructure to assess the annual risk of seasonal influenza, work could improve usability and coverage of health information exchanges, and other initiatives) and ensure inclusion of entities with supporting functions across scales—including private health care systems that provide data or state and local public health laboratories that collect data—in ongoing system development and evaluation.

### RECOMMENDATION 3

The U.S. Department of Health and Human Services should establish an effective and sustainable science-driven leadership and governance structure for the use of SARS-CoV-2 genome sequences in addressing critical national public health and basic science issues, develop a national strategy, and ensure the funding needed for successful execution of the strategy.

- Leaders of this effort must have sufficient authorities and responsibilities to ensure that key issues are identified and prioritized, representative data are generated, and barriers to data sharing are diminished.
- A national strategy for SARS-CoV-2 genome sequences linked to clinical and epidemiological data should be developed that articulates goals, priorities, and a path for achieving them.
- A board with diverse relevant expertise should be established with broad authority to oversee and advise the national strategy for SARS-CoV-2 genome sequences linked to clinical and epidemiological data, and the delivery of actionable data for related investigations.

To read the full report, please visit  
[nationalacademies.org/SARS-CoV-2-Data-Infrastructure-Needs](https://nationalacademies.org/SARS-CoV-2-Data-Infrastructure-Needs)

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

[www.nationalacademies.org](https://www.nationalacademies.org)