Plan to Advance Data Innovation

A Report by the

Epidemic Modeling and Forecasting Fast Track Action Committee

Subcommittee on Health Security Threats

Committee on Homeland and National Security

of the

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

February 2022

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development enterprise. A primary objective of the NSTC is to ensure science and technology policy decisions and programs are consistent with the President's stated goals. The NSTC prepares research and development strategies that are coordinated across Federal agencies aimed at accomplishing multiple national goals. The work of the NSTC is organized under committees that oversee subcommittees and working groups focused on different aspects of science and technology. More information is available at http://www.whitehouse.gov/ostp/nstc.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, homeland security, health, foreign relations, the environment, and energy, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget with an annual review and analysis of Federal research and development in budgets, and serves as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the Federal Government. More information is available at http://www.whitehouse.gov/ostp.

About the Epidemic Modeling and Forecasting FTAC

The Epidemic Modeling and Forecasting Fast Track Action Committee (EMF FTAC) was established by the Office of Science and Technology Policy (OSTP) to respond to Executive Order 13994 on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats. The FTAC was tasked with identifying mechanisms to innovate public health data as it relates to epidemiological modeling and forecasting for high-consequence public health threats, and was co-chaired by the OSTP, the National Security Council (NSC), and the Centers for Disease Control and Prevention (CDC). Recognizing that data underpin local, national, and international responses to public health threats, the goal of the FTAC was to create a Data Innovation Plan to improve public health data by fostering innovation while expanding the utility of existing data sources.

Copyright Information

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

Chair

Eric Lander, Director, Office of Science and Technology Policy

Executive Director

Kei Koizumi, Acting Executive Director, National Science and Technology Council

EPIDEMIC MODELING AND FORECASTING FAST TRACK ACTION COMMITTEE

Co-Chairs

Greg Hebner, Assistant Director for Intelligence Programs, OSTP

Michael Schmoyer, Assistant Director for Health Security Threats, OSTP

Michael Johansson, Senior Advisor for Infectious Disease Modeling and Analytics, CDC

Maureen Bartee, Senior Advisor and Director for Emerging Biological Threats, NSC

Members

Executive Office of the President

National Intelligence Council

National Security Council

Office of Science and Technology Policy

U.S. Digital Service

Departments

Department of Agriculture

Department of Defense

Department of Energy

Department of Health and Human Services

Department of Homeland Security

Agencies

Agency for International Development

Armed Forces Health Surveillance Division

Assistant Secretary for Preparedness and Response

Biomedical Advanced Research and Development Authority

Centers for Disease Control and Prevention

Central Intelligence Agency

Defense Advanced Research Projects Agency

Defense Health Agency

Defense Threat Reduction Agency

Environmental Protection Agency

Food and Drug Administration

National Center for Medical Intelligence

National Defense University

National Institutes of Health

National Science Foundation

United States Geological Survey

United States Public Health Service

Acknowledgements

The members of the EMF FTAC would like to acknowledge the staff of the Science and Technology Policy Institute, particularly the efforts of Dr. Carly Cox, Dr. Ian Simon, Dr. Lisa Van Pay, Dr. Katherine Fisher, Lara Rubinyi, James Karabin, and Amana Abdurrezak.

Table of Contents

About the National Science and Technology Council i
Table of Contents iv
Abbreviations and Acronyms v
Executive Summary 1
Introduction 3
Principles Guiding the FTAC Plan Development 4
Goals and Vision 4
1. Data Inputs and Acquisition 5
Objectives 6
2. Data Management 7
Objectives 7
3. Data Use and Analysis 8
Objectives 8
4. Data Communication and Outputs 9
Objectives 9
5. Cross-cutting Themes and Policy Recommendations 10
Recommendations 10
Conclusion 13

Abbreviations and Acronyms

CDC	Centers for Disease Control and Prevention
COVID-19	coronavirus disease 2019
EO	Executive Order
FTAC	Fast Track Action Committee
HHS	Department of Health and Human Services
NIH	National Institutes of Health
NSC	National Security Council
NSF	National Science Foundation
NSTC	National Science and Technology Council
OSTP	Office of Science and Technology Policy

Executive Summary

Executive Order 13994 on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats calls for development of an improved public health infrastructure to effectively prevent, detect, and respond to future biological threats. Section 4 of the Executive order tasks the White House Office of Science and Technology Policy to "develop a plan for advancing innovation in public health data and analytics in the United States." The Office of Science and Technology Policy formed a National Science and Technology Council Fast Track Action Committee to develop this plan to enhance data innovation and to ensure that epidemiological modeling and forecasting can support preparedness for and response to high-consequence biological threats, in particular respiratory pathogens. This plan also encourages the development of new and innovative thinking about data sources and their applications, and identifies mechanisms for data innovation that should be applied to public health data needs outside of epidemiological modeling, such as needs in maternal health, mental health, veterans health, and other areas.

The objectives and policy recommendations in this plan were designed to guide the new National Center for Epidemic Forecasting and Outbreak Analytics and support contributors and users of robust data sources across multiple sectors. The plan is structured around a four-element data life cycle framework composed of data inputs and acquisition, data management, data use and analysis, and data outputs, including interpretation and communication.

Elements and Objectives

1. *Data Inputs and Acquisition* is broadly defined to include data, expertise, and processes necessary to create the building blocks for epidemiological modeling and forecasting. Objectives in this element include: identify and assemble interdisciplinary teams to support epidemiological modeling and forecasting; develop mechanisms to improve data collection and ensure data quality; establish a baseline database of critical data; and develop a process to encourage innovation, enhance data sharing, and improve versatility.
2. *Data Management* encompasses the storage and documentation of data in a way that allows the datasets to be identified, accessed, and understood by all users. Objectives in this element include: build infrastructure for data storage, management, and access; establish a flexible metadata format that supports data discovery, indexing, and cataloguing for diverse data types; support research to develop modern data standards and keep them updated; and support research to develop tools to maintain data.
3. *Data Use and Analysis* involves efforts to generate actionable insights from data, including analysis, early warning, modeling, and forecasting. Objectives in this element include: facilitate the integration of data from multiple sources and disciplines; develop infrastructure to facilitate the use and analysis of data; and formalize mechanisms to develop and validate models and forecasts through interdisciplinary teams.
4. *Data Communication and Outputs* includes standardized methods to communicate forecasting results, platforms and personnel to aid in interpretation of information, and dashboards that link data, model outputs, and key messages. Objectives to drive innovation in this element include: support research to identify effective practices for communicating results and model limitations; develop and disseminate standards for communicating model results to key audiences; develop integrated dashboards where users can access data and models in real-time; identify, train, and organize professional communicators to interpret and disseminate model results; and develop forecast and modeling short-courses and trainings.

We also identify four key themes that cut across all stages of the data life cycle outlined in this Data Innovation Plan and which are necessary to enhance the utility of public health data in epidemiological modeling and forecasting. These four themes serve as the framework for policy recommendations.

1. *Enable workforce development and readiness*: A pipeline to train and sustain a workforce with modeling capabilities, and empower them to better communicate the results and implications of epidemiological modeling is needed in support of data innovation. To facilitate this, the assessment, prioritization, and implementation of current or new hiring authorities; robust and ongoing training programs for data collection, management, and use; and support for a diverse pool of government and non-government experts to assist in the response to high-consequence public health threats are needed.
2. *Build critical infrastructure*: Infrastructure that supports rapid collection, management, analysis, and communication of public health data across jurisdictions and that can drive innovation in epidemiological modeling and forecasting is needed. To build this infrastructure, a unified system for finding, accessing, storing, and using data is needed. An interagency working group is also needed to assess how current authorities to enable and incentivize collection, reporting, and use of critical public health information is deficient and could be improved.
3. *Develop partnerships*: Partnerships are needed across each stage of the data life cycle to advance innovation in public health data and respond to high-consequence public health threats. To develop and sustain these partnerships, the U.S. Government should utilize an interagency working group to develop data standards; facilitate partnerships to enable international data sharing, leverage private sector resources, and promote coordination across sectors; incorporate feedback mechanisms during each stage of the data life cycle; and update transaction authorities.
4. *Support continuous innovation*: Mechanisms to support innovative research across all stages of the data life cycle should be established to build capacity and capabilities in epidemiological modeling and forecasting. Specifically, a dedicated entity, such as an interagency body, should identify key challenges and opportunities for collaborative, cross-sector research partnerships; joint solicitations could be leveraged to address key challenges at the intersection of multiple fields and foster closer collaboration and partnerships with academia and the private sector; and efforts to address health inequities resulting from high-consequence public health threats should be strengthened and prioritized.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic revealed strengths and weaknesses in America's ability to respond effectively to infectious disease outbreaks and to the risks posed by future accidental or intentional release of biological pathogens. To maximize America's ability to respond to future pandemic-level health challenges, the Nation needs to reinforce its existing strengths while developing greater capacity and capability in public health data management and use.

Executive Order 13994 on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats (Data EO)[1] calls for the creation of an improved public health infrastructure to effectively prevent, detect, and respond to future biological threats. Recognizing the critical role of data in responding to the emergence of pandemic-type events, Section 4 of the Data EO tasks the White House Office of Science and Technology Policy (OSTP) to "develop a plan for advancing innovation in public health data and analytics in the United States." This plan is intended to synchronize with Federal efforts of the COVID-19 Task Force, the American Pandemic Preparedness Plan,[2] Department of Health and Human Services (HHS) and Centers for Disease Control and Prevention (CDC) pandemic-response activities, and others. OSTP established a National Science and Technology Council (NSTC) Fast Track Action Committee (FTAC) to develop this plan, which applies U.S. strengths as an innovator to the challenge of harnessing and expanding data that will improve health security and public health during all stages of high-consequence events.

Activities in this plan support related Federal efforts, including efforts to revitalize biosurveillance capabilities, establish a National Center for Epidemic Forecasting and Outbreak Analytics,[3] and identify epidemiological modeling and forecasting needs as described at the November 2020 National Summit on the Science and Technology of Epidemiological Modeling and Prediction.[4] This plan will drive the development of new and innovative thinking about data sources and their applications in an increasingly cyber-connected world while aligning with the overarching vision of the American Pandemic Preparedness Plan,[5,6] and identifies mechanisms for data innovation that should be applied to public health data needs outside of epidemiological modeling, such as needs in maternal health, mental health, veterans health, and other areas. As such, this plan capitalizes on the current momentum to respond to and mitigate future pandemic events by leveraging related activities to ensure impact.

---

[1] https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/21/executive-order-ensuring-a-data-driven-response-to-covid-19-and-future-high-consequence-public-health-threats/

[2] https://www.whitehouse.gov/wp-content/uploads/2021/09/American-Pandemic-Preparedness-Transforming-Our-Capabilities-Final-For-Web.pdf

[3] https://www.whitehouse.gov/briefing-room/statements-releases/2021/01/21/national-security-directive-united-states-global-leadership-to-strengthen-the-international-covid-19-response-and-to-advance-global-health-security-and-biological-preparedness/

[4] https://www.ida.org/-/media/feature/publications/s/su/summary-report-of-the-national-summit-on-the-science-and-technology-of-epidemiological-modeling/d-21545.ashx

[5] https://www.whitehouse.gov/wp-content/uploads/2021/09/American-Pandemic-Preparedness-Transforming-Our-Capabilities-Final-For-Web.pdf

[6] Specifically, this plan aligns with efforts in the American Pandemic Preparedness Plan related to monitoring the spread and evolution of high-consequence public health threats and strengthening real-time analytics and model development.

Principles Guiding the FTAC Plan Development

- Infectious diseases do not respect borders. Coordination and collaboration among Federal, local, State, Tribal, and global authorities, in addition to non-governmental partners, is needed to identify, mitigate, and respond to high-consequence public health threats. Data and new sources of information that encompass many public and private sectors, data and security solutions, and global developments are needed to support the response to high-consequence biological threats and mitigate the effect upon human health, livelihoods, and economies.
- The COVID-19 pandemic has had a disproportionate effect on many of the Nation's most vulnerable communities and made apparent social and health inequities by race, ethnicity, geography, disability, sexual orientation, gender identity, and other factors. The Administration is committed to addressing these inequities and preparing for future high-consequence public health threats.
- Although the COVID-19 pandemic has had devastating effects on human health and livelihoods, future pandemics may be substantially different than COVID-19, occur with increasing frequency, and have larger impacts. The United States must explore new and more innovative methods to gather and use data to predict the early emergence and possible trajectory of future viral threats to guide actions during the critical early days. Respiratory pathogens are of particular concern due to their high transmissibility and ability to quickly cause large numbers of casualties, overwhelm health care systems, and weaken economies.

Goals and Vision

Data are foundational to the development of sound policy and effective responses to national security threats, including the threats to economic, military, and societal wellbeing from future high-consequence health events. However, the COVID-19 pandemic has made apparent the need to improve the collection and use of public health data to develop elements of a national response plan. The goal of this Data Innovation Plan is to reinforce and improve existing health-related data sources to ensure that epidemiological forecasting, modeling, and predictive analytics can support preparedness for and response to high-consequence biological threats, in particular, respiratory pathogens.

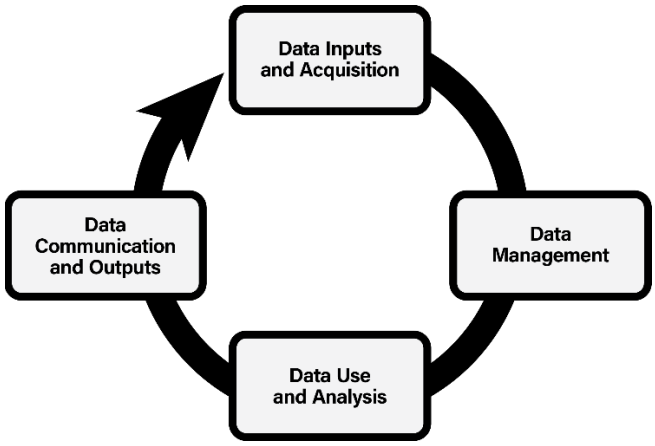

Figure 1. Data Life cycle Framework. The Data Innovation Plan is structured around a data life cycle with four elements: 1) data inputs and acquisition; 2) data management; 3) data use and analysis; and 4) data communication and outputs. Although listed sequentially, these elements are interconnected and expected to evolve over time.

The Data Innovation Plan is structured around a data life cycle framework (Figure 1) to inform the development of a holistic environment where innovation in public health-related data is fostered and enhanced. This plan uses a four-element framework composed of data inputs, data management, data innovation and analysis, and data outputs, including interpretation and communication. The four elements of the framework also have cross-cutting elements, which will be discussed later in the plan. Specifically, the four elements of the framework are:

1. **Data inputs and acquisition** builds a knowledge foundation through identification, collection, and acquisition of information that underpin models. Data inputs also include coordination of subject matter experts in advance of pandemic events;
2. **Data management** includes the curation, sharing, storage, and archiving of data;
3. **Data use and analysis** includes efforts that use, integrate, model, or analyze data, including the exploration of new data sources and methods; and
4. **Data communication and outputs** focuses upon interpreting results and communicating and disseminating information gleaned from the data, models, or forecasts.

Building on the data life cycle framework, this Data Innovation Plan provides an overarching vision for a future environment where:

- Data are identified, collected, stored, archived, and integrated from diverse sources and are sufficiently granular to provide insight on multiple facets of epidemics such as assessing differential vulnerability to severe infections across different factors (e.g., ages, races, and ethnicities), identifying environmental components of transmission risk, or forecasting trends and health system resource needs;
- Data are continually updated and securely accessible to users and researchers generating models and forecasts for high-consequence public health threats through mechanisms that protect patient privacy;
- Results from models and forecasts are effectively interpreted and communicated in a timely manner to all relevant groups from a central source that provides data as locally as feasible;
- Facilities and organizational structures support data management and use, coordination and collaboration across agencies, organizations, industries, and governments; and
- Basic and applied research in data, modeling, and communication continue to drive innovation and enhance the Nation's ability to effectively respond to outbreaks and pandemics.

The following sections expand on the overarching vision and provide detailed objectives of each of these elements.

These objectives and later policy recommendations are intended to guide and support the new National Center for Epidemic Forecasting and Outbreak Analytics and the many U.S. Government agencies and partners in State and local governments, academia, and the private sector that are both contributors to and users of robust data and analyses. A common theme across the FTAC working groups was the need for interagency coordination and cooperation to meet the shared threat from future high-consequence health events.

1. Data Inputs and Acquisition

Data inputs and acquisition is broadly defined to include data, expertise, and processes necessary to create the data foundation for epidemiological modeling and forecasting. Data inputs include the generation of critical core databases required to assess baseline health and health system status

and to develop epidemiological models and forecasts that can detect deviations that may be indicative of the trajectory of possibly high-consequence health events. Data inputs also include processes to collect and assess new data sources, such as data sharing agreements across agencies, nations, and sectors to allow data to be rapidly collected, shared, utilized, and updated in times of need. This element will create an environment that is open to exploring and generating new and novel data sources that may demonstrate utility to predict the trajectory of future health threats. Further, data utility can be greatly enhanced by the increased use of subject matter experts and data stewards to respond to the evolving needs of users and ensure that all the data are vetted, accessible, and useful.

Objectives

1.1 **Identify and assemble interdisciplinary teams to support epidemiological modeling and forecasting**. Diverse teams of subject matter experts and data stewards[7] that can leverage information and data from many fields to inform models and forecasts and add their considerable domain experience are needed to support the response to high-consequence events. Individuals in these teams should be identified in advance, and mechanisms should be developed to rapidly assemble and support these teams in times of need. Team membership should be periodically reviewed to include additional expertise as outbreaks evolve. Teams should also exercise their capabilities and develop methods to maintain relationships in-between outbreaks.

1.2 **Ensure data quality.** Continuous efforts are needed to identify and assess the representativeness, quality, granularity, and accuracy of public health data for use in epidemiological modeling. Approaches for assessing and quantifying the data eligible for inclusion, inaccuracies, and biases of different data streams (e.g., representativeness) should be developed to ensure that the data used in models and forecasts are as high quality as possible, uncertainty is articulated, and that the intrinsic limitations of datasets are clear to users throughout the data life cycle.

1.3 **Establish a baseline database of critical data.** In conjunction with 1.2, public health policy makers, epidemiologists, data scientists, Electronic Health Record vendors, and healthcare workers (1.1) should collaborate on what is needed to establish baseline data that can be utilized to indicate the possible emergence of health threats and forecast its early trajectory. Once established, a one-stop-shop for baseline data should be created. This database could be used to calculate baseline prevalence of specific health conditions, health system status, health-related behavior, and vulnerability indices of geographical areas or populations. These datasets could directly inform resource allocation, monitor status, and assess risks in the early response to an emerging health threat, as well as enable modeling and forecasting of those components. Baseline datasets should have a clear organizational data steward that can act as a point of contact and understand the data provenance, be reviewed regularly, be updated to respond to the changing public health landscape, and include the flexibility to take advantage of new data sources.

1.4 **Develop a culture and processes to encourage innovation, enhance data sharing, and improve versatility.** Create and foster an environment that is open to generating and evaluating new data sources that could be combined with or complement existing data to provide unique insights into the likely trajectory of future health threats. Standardized mechanisms to access data, such as through data use agreements, cross-sector partnerships, and incentives should be

[7] Data stewards include individuals who can understand dataset origins and nuances and can help inform appropriate data use.

considered across agencies and sources to allow rapid, transparent, and secure integration of and access to new and existing data from diverse sources. Since some data may not be widely available due to privacy or security concerns, mechanisms to share that data securely should be established a priori. The data input process should be flexible and able to quickly incorporate new sources of data, enabling fluid sharing with transparent data use agreements across governments, academia, the private sector, international entities, and others. It should also be responsive to stakeholder feedback and adaptable to advances in the field.

2. Data Management

Proper data management, including transparency and protection, is foundational to the future use of collected data. Data management encompasses the storage and documentation of data in a way that allows the datasets to be identified, accessed, and understood by all users. This section will address the necessary components for modernized data management to create an environment that will enable efficient data sharing and use within a protected structure. Further, it is anticipated that the unprecedented aggregation of data proposed in this plan could produce positive benefits for citizens and commercial applications that are currently unimagined.

Objectives

2.1 **Create infrastructure for data storage, management, access, and protection.** Utilizing the process developed in 1.4, infrastructure should be developed to support the storage, management, and protection of data from diverse sources. This infrastructure should enable all contributors and users to submit or access data in accordance with transparent access controls, further enabling data sharing and creating a data environment that is as open as possible, which is critical to promoting innovation. System flexibility will allow data to be transparently updated with version tracking and will allow unanticipated, transformative applications to be developed as data and capabilities evolve. The infrastructure should also support international engagement with approved collaborators and be highly accessible even in the event of global disruptions.

2.2 **Establish a flexible metadata format that supports data discovery, indexing, and cataloguing for diverse data types.** Metadata that enable end users to understand the meaning and sources of datasets are critical to maximize their use. Agencies, members of the private sector, academic institutions, and other entities that collect data should be engaged to develop metadata standards in a format that enables discovery and indexing across all datasets and promotes standard uptake. Efforts to identify the minimal elements or characteristics of data necessary to maximize their use are also needed. These metadata should include: originating organization and data source, date of collection, collection methodology, and any data processing, etc. Data stewards can support improved data use, both while these standards are being developed and after standard implementation. Metadata standards will also improve security and data protection.

2.3 **Support research to develop modern data standards and keep them updated.** Research should be supported to identify, and develop as needed, agile standards for data collection, protection, quality, access, and management. Agencies, members of the private sector, academic institutions, and other entities that collect data should be engaged to develop data standards that support and promote data collection and use through the data life cycle. Standards should be applied to new data streams as well as existing ones.

2.4 **Support research to develop tools to maintain data.** These tools should apply to data types that are tracked over time as well as data and models that are continuously updated. Tools should provide mechanisms for transparently archiving older versions and tracking the provenance of all updates. All tools should be scalable and able to incorporate additional data as they emerge.

3. Data Use and Analysis

Data use and analysis includes efforts to generate actionable insights from data, including analysis, modeling, and forecasting. Improved data analysis requires data availability with clear metadata and use policies as well as an infrastructure that enables agile data use and collaboration. At this stage, experts across disciplines must work together to develop innovative solutions to many of the challenges faced in improving public health data and analytics. For example, potential privacy and transparency concerns, partially due to the aggregation of vast datasets, must also be addressed. In addition to privacy issues, factors that contribute to social and health inequities need to be considered. Reducing barriers to data analysis combined with support for development of analytical tools will enable greater innovation in data analysis, modeling, and forecasting to support the response to public health events.

Objectives

3.1 **Facilitate the integration of data from multiple sources and disciplines.** Most data are collected and managed at the local level, and differences across these datasets make integration non-trivial. Numerous barriers make it challenging to integrate datasets outside of traditional public health fields, such as the private sector. Objectives listed in previous sections, including actions to expand data sharing agreements and encourage system flexibility (1.4), as well as data standardization (2.3), will be integral to improving efforts to integrate datasets.

While integration of datasets will improve our ability to more accurately monitor and model dynamic health threats and better predict the effectiveness of interventions and mitigation proposals, it also increases the amount of health information that could potentially be aggregated, raising privacy concerns. Therefore, in addition to improved methods to integrate datasets, approaches are needed to support privacy-preserving record linkage, to maximize the utility of data while preserving privacy. Partnerships across sectors and efforts to develop innovative solutions to address privacy concerns are key aspects of this objective. Integration of data and analysis should also include sensitive, controlled, or classified government information to provide decision makers with a complete understanding.

3.2 **Develop infrastructure to facilitate the use and analysis of data.** In addition to the infrastructure needs described in 2.1, infrastructure is needed to support the development of platforms for modeling and forecasting. A unified system could include interfaces that allow for a range of activities to support actions across the data life cycle, including flexibility to ingest and export new data sources, the ability to run internally or externally developed interoperable code or models, version-controlled model code and outputs, and output to or incorporation of dashboards (further discussed in 4.3). Design considerations for the system could also include an organizational structure that permits modeling units to scale up with breadth and depth as required.

3.3 **Formalize mechanisms to develop and validate models and forecasts through interdisciplinary teams.** Mechanisms to design models that innovate on and strengthen existing practices are needed and should be established through robust teams and partnerships. These teams

can be charged with developing both processes and products to advance interoperable and scalable models, interdisciplinary models, and systems for validation. For example, models that utilize both traditional and non-traditional public health data, as well as incorporate other factors that play a role in social and health inequities, can be used to generate more reliable forecasts and help inform community-specific interventions. In this context, partnerships between data generators, model developers, model users, including policy makers and other key stakeholders can be leveraged to improve modeling capabilities and outputs.

4. Data Communication and Outputs

Data communication and outputs are critical to ensure that data and model results reach decision makers and the public in a timely and actionable manner, and are appropriately communicated for use in decision making. Important components of data outputs include standardized methods to communicate forecasting results, platforms and personnel to enable access to and interpretation of information, and dashboards that link data, model outputs, and key messages.

A key part of communications is not only sharing the current status, but also the likely and unlikely implications of the model results on the functioning of government and society. With a central source supporting the transmission of information, communicators can clearly and effectively share epidemiological forecasts in unified and more accessible ways while conveying the nuances in the underlying data sources, model methods, results, uncertainties, and limitations.

Communication should also be evidence-based and unbiased, and conducted in a way that is useful to decision makers. Data communication should be aligned with the best practices in crisis communication, and contribute to the establishment of communication protocols. Crucial resources and personnel are required to ensure effective communication of model results, support decision making, and facilitate a more coordinated response to public health threats, including policy decisions.

Objectives

4.1 **Research and identify effective practices for communicating complex results and model limitations.** Forecast results, including model limitations and uncertainties, must be effectively communicated to ensure they produce actionable insights. Research should be supported to evaluate and improve the communication methods (e.g., written, verbal, visual, multimedia) for relaying public health information to diverse audiences. This research should assess different techniques or combinations of techniques to identify the best approaches to communicate results for action and should continue in the future as new products and communication tools are developed.

4.2 **Develop and disseminate standards for communicating model results to key audiences.** Currently, communication strategies for forecast and modeling results have limited standardization, which can result in ambiguity based on interpretation of the developers and users, including the public and decision makers. Evidence-based reporting and communication standards (leveraging work in 4.1) should be developed in conjunction with feedback from the target audiences (e.g., Federal, State, local, and Tribal governments, community leaders, the public, policy makers) to ensure communication efforts will be consistent, reproducible, and transparent. Collaboration between agencies, the private sector, academic institutions, and other entities that

develop and communicate model results would inform standard development and promote uptake of best practices.

4.3 **Develop integrated dashboards where users can access data and models in real-time.** Leveraging the data and models generated in Sections 1 and 3, integrated dashboards with visualization tools should be developed collaboratively to allow authorized users to access and view underlying data and updated model results in real-time. Dashboards should be able to scale to different geographic resolutions (e.g., State, county), reflect best practices (e.g., 4.2), and should become the official source of model and forecasting results for the U.S. Government via a common platform. Dashboards should be easily understood and accessible to all users.

4.4 **Identify, train, and organize professional communicators to interpret and disseminate model information and results.** These communicators, serving as epidemiological-forecasting "weather persons," should become trusted sources for sharing information. By embedding the communicators within relevant organizations, they can be empowered to communicate results independently to diverse audiences and ensure that messages are as scientifically accurate, accessible, and unbiased as possible. The information should be communicated in a way that facilitates all decision making.

4.5 **Develop forecast and modeling short-courses and trainings.** Courses and trainings should provide an overview of a) the types of public health decisions that modeling and forecasting can best inform, b) methods and validation, c) uncertainty and its interpretation, and d) translation of results into action. These courses will ensure forecast users at all levels (Federal, State, local, and Tribal community leaders, professional communicators, and the public) understand how to use and interpret forecast and model results and have examples of the application of these results to inform decision making.

5. Cross-cutting Themes and Policy Recommendations

Key themes within this Data Innovation Plan cut across all stages of the data life cycle and are critical to enhance the overarching utility of public health data in epidemiological modeling and forecasting to mitigate emergent outbreaks. Four cross-cutting themes emerged that serve as the framework for policy recommendations and that will create an environment where high-consequence public health threats can be mitigated from affecting human health, livelihoods, and economies. These themes are:

1. Develop and maintain a talent pipeline capable of leveraging the data and tools necessary to respond to future high-consequence public health threats;
2. Invest in infrastructure that allows data to be collected, managed, and used easily and securely;
3. Partner across sectors, governments, and nations to facilitate data and tool sharing; and
4. Support research to continuously promote and improve innovation in data collection, management, analysis, interpretation, and communication.

Recommendations

5.1 **Enable workforce development and readiness.** Innovation across the data life cycle requires an agile workforce with the capability to support cutting-edge methods development, implement systems and technology integration, advance modeling capabilities, and better communicate the results and implications of epidemiological modeling. A pipeline to train and recruit individuals

for new types of public health data-related roles such as research and development, technology, operational analysis, strategic oversight, and data interpretation is required. Efforts to improve ongoing training to equip employees with new skills and/or maintain skills needed to work in the field would support the steady development of the workforce and enhance retention. Emphasis should be placed on efforts to recruit and train individuals from diverse backgrounds, and with a broad mix of skills across multiple disciplines. To facilitate this goal:

5.1.1 Assessment, prioritization, and implementation of current or new hiring authorities are needed to recruit and retain data scientists from diverse backgrounds.

5.1.2 Robust and ongoing training programs to equip employees with new skills or maintain skills are needed to create a pipeline of workers who can collect, curate, manage, use, and leverage the flow of data and its interpretation and application.

5.1.3 Support for a diverse pool of government and non-government data, modeling, and forecasting experts, as well as pre-approved public health communicators, who can be accessed rapidly, is needed.

5.2 **Build critical infrastructure.** Infrastructure that reduces barriers and supports rapid collection, management, analysis, and communication of public health data across jurisdictions can drive innovation in epidemiological modeling and forecasting. Data sharing mechanisms should be in place to enable flexible information sharing with and between agencies, State, local, Tribal, and territorial governments, as well as with the private sector and academia for use by modelers during public health emergencies. Flexibility and continuity of operations for all public health-related infrastructure is necessary to ensure a response-ready workforce is in place prior to and during prolonged emergency conditions. Open architectures will encourage future applications of the data archive. To build critical infrastructure:

5.2.1 A unified system for identifying, accessing, collecting, storing, and using data should be established. The system should facilitate connectivity of data systems and analyses across the interagency, to include the ability to access and fuse sensitive government information. Infrastructure should be flexible and able to support continuity of operations at all classification levels during prolonged emergency conditions. Standards should be implemented to facilitate the interoperability of data and models, including the use of big data and high performance computing when appropriate.

5.2.2 Global data sharing mechanisms should be established to support the unified system (e.g., standardized multi-use agreements/standard data use agreements across international, State, Tribal, local, and territorial jurisdictions). When possible, data should be shared publicly to facilitate innovation by public and private sector partners.

5.2.3 An interagency working group should be established to assess how current authorities to enable and incentivize reporting and use of critical public health information from healthcare providers and laboratories is deficient and identify what new or updated authorities are needed. New authorities could include the ability to mandate and/or incentivize data reporting for critical data. Updated authorities could include provisions to share and use de-identified surveillance data and increase agility in novel reporting mechanisms.

5.3 **Develop partnerships.** Partnerships among sectors, nations, and governments are needed across each stage of the data life cycle to advance innovation in public health data and respond to high-consequence public health threats. These partnerships should facilitate national and international data sharing, leverage private sector resources, and promote coordination among Federal, State, local, and Tribal governments. Partnerships should also enable incorporation of

feedback from key stakeholders during each stage of the data life cycle to ensure that data collection, management, use, and output methods are continuously refined and improved. To develop and sustain these partnerships:

5.3.1 An interagency working group should be established and charged with creating a network of relevant agencies, institutions, organizations, private industry leaders, and other stakeholders to review, update, and develop standards for all four elements of the data life cycle. The interagency working group should consider mechanisms to incentivize and promote adoption of these standards by all stakeholders. The network should also coordinate Federal efforts and establish interagency research priorities.

5.3.2 Partnerships should be facilitated through memoranda of understanding or data use agreements to support international data sharing, enable secure and protected discussions, leverage private sector resources, and promote coordination among Federal, State, local, and Tribal governments, especially in times of emergency.

5.3.3 Partnership agreements should incorporate feedback mechanisms during each stage of the data life cycle to enable continuous improvement and ensure that communication works in all directions. The elements of the data life cycle framework must serve the needs of all stakeholders.

5.3.4 Other Transaction Authorities could facilitate the rapid establishment of partnerships to support innovation across the stages of the data life cycle in preparation for and response to critical health threats. An interagency working group should be established to assess how current authorities are deficient and identify modifications.

5.4 **Support continuous innovation.** Mechanisms to support innovative research across all stages of the data life cycle should be established to build capacity and capabilities in epidemiological modeling and forecasting. In particular, support for new data sources and novel methodologies coupled with studies to validate their utility is critical. New data sources offer opportunities to improve precision, timeliness, and representativeness, and to drive integrated, multidisciplinary analyses that leverage diverse epidemiological drivers such as human behavior, socioeconomic, pathogen evolution, environment, and planetary health. This effort should incentivize novel applications of public health data to respond to future pandemics and encourage development of data management methods applicable to information outside of epidemiological modeling. Specifically:

5.4.1 Efforts to address health inequities that may result from high-consequence public health threats should be strengthened and prioritized.

5.4.2 A dedicated entity should identify key challenges and opportunities identified for collaborative, cross-sector research partnerships coupled with support to drive innovation.

5.4.3 Joint solicitations, through entities such as CDC, NIH, and NSF, could be leveraged to address key challenges at the intersection of multiple fields, including but not limited to the integration of social, behavioral, and economic sciences with pandemic modeling and response. These should also establish review mechanisms to validate the utility of proposed new and novel data sources or analysis techniques.

5.4.4 Focused solicitations that foster closer collaboration and partnerships with academia and the private sector to support the use and application of novel data sources should be implemented.

Conclusion

The COVID-19 pandemic has focused the efforts of the country and this Administration on how to more effectively respond to future infectious disease outbreaks, irrespective of their origin. In response to Executive Order 13994 on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats, this FTAC has developed a Data Innovation Plan to harness our strengths as an innovator to ensure the critical data that underpin policy actions and effective response will improve health security and public health during all stages of high-consequence health events. This document highlights multiple key areas within a data life cycle framework for future support such as developing a talent pipeline; building critical infrastructure; developing partnerships to facilitate data sharing and modeling; and investing in basic and applied research and development. Although these efforts will take time and resources, they will serve as the groundwork for efforts at the National Center for Epidemic Forecasting and Outbreak Analytics and other activities benefiting from data innovation. Through these efforts, the use of all data, including those relevant to and beyond the application of epidemiological models and forecasts, can be maximized to promote the health and prosperity of the Nation.