Accenture Federal Services

Building next-generation public health surveillance

How federal agencies can modernize public health surveillance infrastructure, data, and analytics to gain real-time, actionable insight and impact on a national scale

Executive summary

Public health surveillance in the United States spans the monitoring of infectious diseases, chronic diseases, injuries, and mental health conditions, as well as social determinants of health. Surveillance captures data on behavioral risk factors, preventive actions, cases, program or treatment costs, and other factors that are relevant to the cause or spread of disease.

The COVID-19 pandemic cast a spotlight on significant gaps in public health surveillance data infrastructure and methods – including fragmented data, lack of interoperability, analytical approaches that cannot scale, and cumbersome data collection and reporting. In response, the Centers for Disease Control and Prevention (CDC) is leading multiple initiatives to transform our public health surveillance system into one that is **connected, resilient, adaptable, sustainable**, and **response-ready**.

To succeed, CDC and public health data partners must both modernize the data infrastructure and expand data collection and sharing. A **scalable, federated data mesh** would allow agencies to centralize data governance and integrate privacy-preserving record linkage (PPRL) technology to facilitate HIPAA-compliant patient matching – without creating bottlenecks.

Advancing data interoperability is key to expanding the universe of data used to derive public health insights. By driving the adoption of bulk Fast Healthcare Interoperability Resources (FHIR®) standards and employing artificial intelligence, agencies can harmonize existing health data and ingest data from nontraditional sources to inform analysis and predictions related to public health threats.

With next-generation infrastructure and accessible, interoperable data, CDC can establish an analytical pipeline with unprecedented robustness.

For example, this pipeline could fuel **agent-based models and simulations with sufficient power to derive real-time insights** – for better policy and programs focused on prevention, control, and response. By employing **intelligent automation**, agencies can ensure that these modernization efforts do not create additional burdens on already-strained public health workers – effectively doing more with less.

Success will hinge on alignment between public health leaders and federal data partners.

Action plan for alignment

1 **Define one or more discrete, priority use cases** to demonstrate the value of data solutions.

2 **Select data partners** whose data sources can be integrated into a data mesh solution.

3 **Create a participatory governance framework** to address policy, technical, and operational considerations.

Working together, CDC and its data partners can make sound, strategic investments while supporting effective change management.

With a fully modernized surveillance system, public health leaders can draw on real-time data and insights to understand disease burden, predict future risk, develop and evaluate prevention and control strategies, and – ultimately – save lives.

Surveillance has long been the cornerstone of public health practice.

The collection and dissemination of surveillance data informs individuals' health behaviors, public policy, national security, and global economies.[1]

Led by the CDC and implemented by all 50 states and more than 3,000 local jurisdictions and territories,[2] public health surveillance in the United States spans the monitoring of infectious diseases, chronic diseases, injuries, and mental health conditions, as well as social determinants of health. Surveillance can capture data on every aspect relevant to the cause or spread of disease – behavioral risk factors, preventive actions, cases, program or treatment costs, and more.

The COVID-19 pandemic put public health surveillance – and its urgent need for modernized systems and methods – in the spotlight.[3] To identify, contain, and prevent outbreaks, state and local public health agencies undertook the massive task of tracking cases, variants, vaccinations, and hot spots and sharing that data with federal agencies. This was no simple task – for example, Politico reports that "in Washington state, health officials went from tracking 30,000 disease lab reports a month in 2019 to 30,000 a day during certain points in 2020."[4]

This sharp increase in demand strained what were already significant gaps in public health surveillance data infrastructure and methods, including:

- Fragmented data
- Lack of interoperability
- Analytical approaches that cannot scale
- Delayed and cumbersome data collection and reporting

Common sources of public health data

[1] Surveillance (worldbank.org)
[2] Public Health Surveillance and Data: Where does our data come from? (cdc.gov)
[3] Tracking COVID-19: U.S. Public Health Surveillance and Data (congress.gov)
[4] Inside America's Covid-reporting breakdown (politico.com)

Hospitals, healthcare providers, and laboratories use a variety of systems to collect data – some required by law, others on a voluntary basis.

Typically, they report data to state and local public health agencies, which share the information with CDC and other federal agencies. These agencies aggregate, deidentify, synthesize, and disseminate the information to inform policymaking, public awareness, and research – a process that can often take months or years after the data was initially collected.

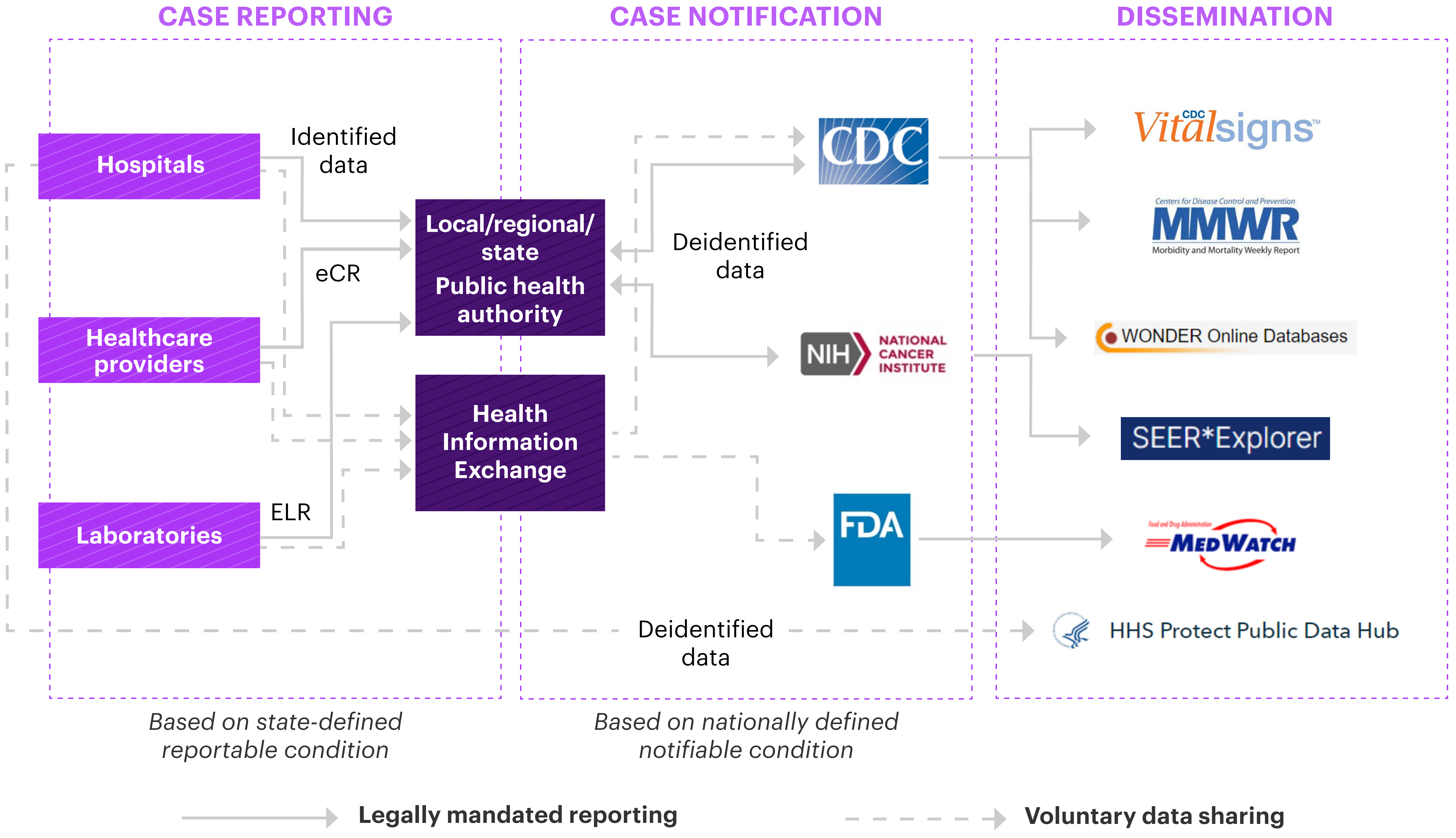
Typical flow of public health data

Many current systems rely on disease-specific monitoring and manual data entry, which places a substantial burden on federal data partners. State and local reports to CDC are often delayed because the systems and data are not interoperable.

CDC encourages standardization, but it lacks the authority to receive data directly without establishing a data use agreement with each state and local jurisdiction. As a result, the agency must manually clean the data before conducting the analyses needed to provide a national, aggregated picture of public health.[5] This can significantly delay the sharing of data with providers and other trusted partners with important roles in public health response.

[5] The Challenges of Integrated Surveillance for National Public Health Institutes during the COVID-19 Pandemic (ianphi.org)

Sample flow of public health data



Fragmented, standalone systems

In addition to CDC's core surveillance activities, more than 100 separate, standalone systems managed by multiple agencies track specific diseases and events.[6] For example, the U.S. Food and Drug Administration (FDA) monitors the safety of regulated medical devices, and the National Cancer Institute tracks cancer trends and statistics. These data are disseminated through agency-specific reporting channels and, in some cases, made available for research in data hubs.

With more modernized data infrastructure, public health leaders can better identify and contain outbreaks, understand disease burden, guide policy changes, evaluate and improve prevention and control strategies, and target research investment where needed.

CDC core surveillance activities

Program	Focus
National Notifiable Disease Surveillance System (NNDSS)	Infectious and noninfectious diseases
National Syndromic Surveillance Program (NSSP)	Symptoms of patients treated in emergency departments
National Vital Statistics System (NVSS)	Birth and death data

[6] Public Health Surveillance: Preparing for the Future (cdc.gov)

Current efforts to modernize public health surveillance

The United States has made major advancements in notifiable disease reporting, syndromic surveillance, mortality reporting, and electronic lab reporting in the past decade. Building on these efforts, CDC recently launched a comprehensive Data Modernization Initiative (DMI) and a dedicated Center for Forecasting and Outbreak Analytics (CFA). These initiatives are many years in the making and, taken together, they are leading the charge to transform our public health surveillance system into one that is connected, resilient, adaptable, sustainable, and response-ready.

In its first year, DMI achieved significant advancements in real-time data collection, cloud-enabled services, and automation across its core systems.[7] While still in its pre-launch phase, CFA successfully predicted the COVID-19 Omicron variant surge and has since produced models and analyses as part of our nation's response to outbreaks of polio and monkeypox.[8]

CDC's efforts are bolstered by the Office of the National Coordinator for Health Information Technology (ONC)'s work to define standards and practices for interoperable data sharing and inform the incentives driving their adoption. Chief among ONC's accomplishments: Advancing the Fast Healthcare Interoperability Resources (FHIR®) standard and publishing the Trusted Exchange Framework and Common Agreement (TEFCA) to establish a universal floor for interoperability across the country.[9]

While the focus of ONC's efforts to date has been care coordination, attention is shifting to the need for bi-directional exchange with public health agencies and elimination of one-off connections to public health reporting systems.[10]

[7] DMI Snapshot 2021 (cdc.gov)
[8] Early CFA Successes (cdc.gov)
[9] Trusted Exchange Framework and Common Agreement (healthit.gov)
[10] The Nationwide Network Based on the Common Agreement: Benefits for State Governments and Public Health (sequoiaproject.org)

A holistic strategy for making next-generation surveillance a reality

As our nation defines and implements the next round of investments to modernize public health surveillance, agency leaders need a holistic strategy and an unwavering focus on the end goal. Defining and implementing a solution for real-time, actionable data and rapid, accurate insights will require a massive acceleration of efforts across lead agencies and data partners.

As they advance public health systems, agencies will need to simultaneously expand, coordinate, standardize, and streamline data collection and sharing. They can do so by adopting a **scalable, federated data mesh infrastructure** and **further expanding data interoperability**. With a stronger technological foundation and a greater volume of usable data, agencies can then deploy powerful analytical tools at scale that can provide a comprehensive, decision-ready picture of a given public health threat or situation.

At the same time, public health agencies must pursue intelligent automation tools to ensure that the benefits of surveillance modernization do not create additional burdens on already-strained public health workers.

A scalable, federated data infrastructure

Our nation's existing network of siloed, disease-specific systems creates significant redundancies and inefficiencies and – equally important – cannot scale to support the level of data aggregation and access that public health agencies need.

To meet the demands of a modern public health data ecosystem, federal agencies need a scalable, federated data mesh.

By leaving data ownership decentralized, a data mesh allows those who are most knowledgeable to control their data. In a public health context, this means health agencies, insurers, academic partners, and others act as nodes in a network.

Rather than reporting directly to CDC, state and local agencies would make their data products – EHR data, laboratory reports, genomic sequencing information, immunization records, etc. – available via the mesh.

Using a self-service platform powered by robust metadata, search features, and a data catalog, authorized data consumers can find, access, aggregate, and analyze the data. They can also access pre-built algorithms and create new data products and reusable algorithms.

CDC would serve a crucial governance and stewardship role – developing and enforcing implementation guidelines and standards, establishing a data catalog, and executing a privacy layer. Using a privacy-preserving record linkage (PPRL) technology, the privacy layer would maintain HIPAA compliance by enabling patient matching even with deidentified data. For example, PPRL employs hashing to convert names, birthdates, and addresses into encrypted tokens that preserve the original values.

CDC currently has an initiative underway that employs PPRL to further public health and research priorities related to COVID-19. Linking data at the patient level gives a comprehensive view of a person's health, allowing researchers to answer questions that would otherwise require extensive primary data collection or complicated data use agreements.[11]

By operationalizing PPRL with standardized FHIR data components, public health agencies would be able to ingest and collect data from multiple sources and feed those data into scalable analytics and modeling tools.

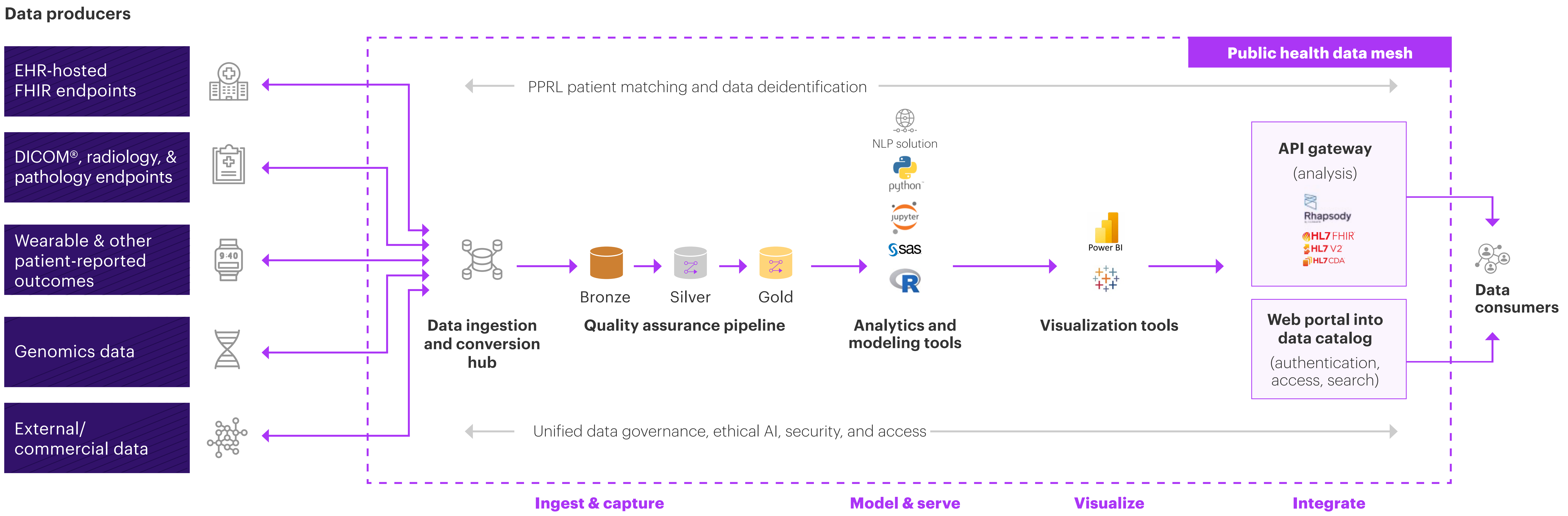
Data mesh vs. data lake

A data mesh can deliver the benefits of a centralized platform or data lake without the potential for bottlenecks:

- Architecture is **cost-efficient, flexible, and scalable**.
- Universal interoperability and privacy layer enforces **data governance standards**.
- Self-service interface **streamlines access and accelerates insights**.
- Decentralized data ownership allows **data-driven innovation**.

[11] HealthVerity awarded CDC contract for real-world healthcare data to advance COVID-19 response (prnewswire.com)

Federating data with a modern data mesh powered by PPRL on FHIR



With appropriate governance, a data mesh would provide access to analysis-ready data products, eliminating the bottlenecks typically associated with centralized reporting and dissemination. As a result, public health agencies could accelerate data aggregation and analysis – and public warnings and outreach – which is particularly critical for fast-moving threats such as infectious diseases.

Data interoperability: shifting the paradigm from data push to data pull

However, data infrastructure is only as successful as the volume and quality of inputs that feed into it. Achieving America's public health goals hinges on widespread adoption of application programming interface (API)-based data standards to accommodate the data volumes necessary for rapid digital reporting in a scalable way.

To that end, public health agencies, surveillance programs, and health information exchanges (HIEs) and their network participants must continue progress toward full adoption of FHIR – and specifically, its RESTful API functionalities such as Bulk FHIR.[12]

With FHIR and Bulk FHIR-enabled APIs, public health agencies could shift from a "push" paradigm that relies on providers to send data. Instead, agencies could adopt a query or subscription-based model ("pull" paradigm) to receive automated case updates.

Aligning data elements to a universal standard

USCDI health data

EHR – FHIR US Core
- Clinical
- Administrative

SDOH – FHIR US Core
- Economic stability
- Education access & quality
- Healthcare access & quality
- Neighborhood & built environment
- Social & community context

Health data outside USCDI

Claims data (X12)
- Inpatient hospitalizations
- Outpatient medical encounters
- Utilization & payment
- Prescription costs

Electronic case reporting – eICR in CDA
- Date of report
- Provider information
- Facility information
- Patient information
- Preferred language
- Occupation
- Health information
- Social & community context

Electronic lab reporting – HL7 v2 or CDA
- Lab test information
- Patient information
- Provider information
- "Ask at entry" data
- Pathogen genomics

Patient registries – Custom
- Cancer
- Chronic diseases
- Rare diseases
- Pregnancy
- Other

Nontraditional data

Other data sources (format varies)
- Google search data
- Social media data
- Mobility data
- Remote sensing data (e.g. wastewater, satellite imagery)
- Other consumer data (e.g. purchase records)
- Wearable data
- Patient genomic data
- Pathogen genomic data

[12] Bulk Data Access Implementation Guide (hl7.org)

Expanding interoperability with intelligent systems

Currently, only EHR data and social determinants of health (SDOHs) are interoperable via the established standard – aka the United States Core Data for Interoperability (USCDI). These data can and should be augmented by structured health data siloed in other agency systems, as well as data from other, relevant sources, including:

- **Geospatial data** such as walkability and access to care
- **Remote-sensing data** such as wastewater testing and satellite imagery
- **Mobility data** from smartphones, GPS, and sensors along highways and roads

By layering additional data from currently siloed health systems and non-health sources, public health agencies can enrich the baseline USCDI data for truly robust insights. Recent efforts have demonstrated the value of multilayered data to track the spread of COVID-19,[13] understand the effects of social distancing,[14] and predict obesity rates,[15] for example.

These results are encouraging but limited in scope. The lack of interoperability across data sources makes it impossible to scale such approaches for real-time, actionable surveillance. While ONC continues to advance and expand USCDI in collaboration with CDC and other stakeholders, this process is incremental by design. In the meantime, CDC must pursue alternate approaches to bring more data into public health models and simulations.

Machine learning feature stores have strong potential to fill in the gaps. This novel tool provides the flexibility required to ingest data – via direct connection or high-throughput API – from sources that use varying data standards. Once ingested, a ML feature store can harmonize that data with FHIR, making it usable in public health models and simulations.

By extending interoperability and connecting the universe of rich, relevant data, public health agencies can boost the accuracy of prevalence estimates, counter-balance biases in traditional data collection, effectively target control and prevention strategies, and better allocate resources.

Right-sizing and ensuring FAIRness of data solutions

To optimize data solutions, it's important to define what's "good enough" to address the problem at hand. For example, generating insights to address obesity requires rich, robust data. In contrast, tracking and forecasting disease outbreaks requires real-time data and rapid analysis.

Similarly, data solutions should follow best practices such as the FAIR guiding principles – which help ensure that data are **F**indable, **A**ccessible, **I**nteroperable, and **R**euseable – for scientific data management and stewardship.[16]

[13] National Wastewater Surveillance System (cdc.gov)
[14] Quantifying human mobility behaviour changes during the COVID-19 outbreak in the United States (nature.com)
[15] Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity (jamanetwork.com)
[16] FAIR Principles (go-fair.org)

Unleashing the potential of a modern data infrastructure

With a federated data mesh infrastructure that allows access to high volumes of rich, interoperable data, a modernized public health surveillance system can deploy advanced analytics and novel technologies to optimize efficiency – all at sufficient scale to produce accurate, real-time insights.

Using natural language processing to analyze complex, unstructured data

A tremendous volume of valuable health data is buried in imaging files, lab reports, and clinical notes. Relatively recent advances in natural language processing (NLP) make it possible to analyze these types of unstructured data.

NLP enables computer systems to understand and interpret human language through topic modeling, sentiment analysis, and other techniques. By capturing complex linguistic relationships, NLP goes well beyond keyword searches to identify common themes or attitudes towards a particular topic from medical record notes, as well as social media data and other large, unstructured data sets.

In recent years, the performance of NLP has improved significantly through what's known as transfer learning – that is, taking a well-honed model and using it to train a new model for a related task. Massive pre-trained language models such as Google's BERT[17] and OpenAI's GPT-3[18] are driving the state of the art across the full range of NLP's capabilities, enabling the development of more powerful models with less training data and computing resources.

To date, public health researchers have successfully employed NLP models to monitor flu-like symptoms mentioned on Twitter, identify public sentiment related to the COVID-19, and pursue other exciting studies.[19] These applications only begin to scratch the surface of NLP's potential – particularly when combined with a federated data infrastructure and extended interoperability – to revolutionize how public health surveillance is conducted on a national scale.

Large-scale modeling for robust, scenario-based insights

Agent-based modeling (ABM) is a computational method for simulating actions and interactions between people and their environment. Public health researchers use ABM to model disease transmission, social influences on health, health behavior outcomes, and evaluate the efficacy of interventions.[20]

The utility of ABM depends on how well the environment and rules that govern agent behavior are understood. With more and better data, ABM simulations can be used to model increasingly complex scenarios.

For example, public health officials could:

- Examine the impact of immunization and introduction of new variants on community spread
- Identify at-risk populations
- Detect hotspots and conditions that promote the spread of the disease
- Proactively evaluate the efficacy and impact of prevention and control strategies

Powered by sufficiently rich data such as demographics, social determinants, vaccination status, geographic and other environmental data, sophisticated agent-based models can predict risk and outcomes, allowing agencies to effectively allocate resources in the interest of public health.

17 Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing (ai.googleblog.com)
18 GPT-3 Powers the Next Generation of Apps (openai.com)
19 Challenges and opportunities for public health made possible by advances in natural language processing (nih.gov)
20 Agent-Based Modeling in Public Health: Current Applications and Future Direction (nih.gov)

Reducing the burden on public health workers with intelligent automation

Greater data collection and more advanced analysis is crucial to furthering our understanding of – and therefore improving – public health. However, surveillance modernization efforts cannot become another burden on the public health workforce. Public health agencies at all levels already face a dire shortage of workers, with roughly 44 percent considering leaving their jobs within the next five years.[21] This makes the adoption of tools such as intelligent automation (IA) an essential step in this journey.

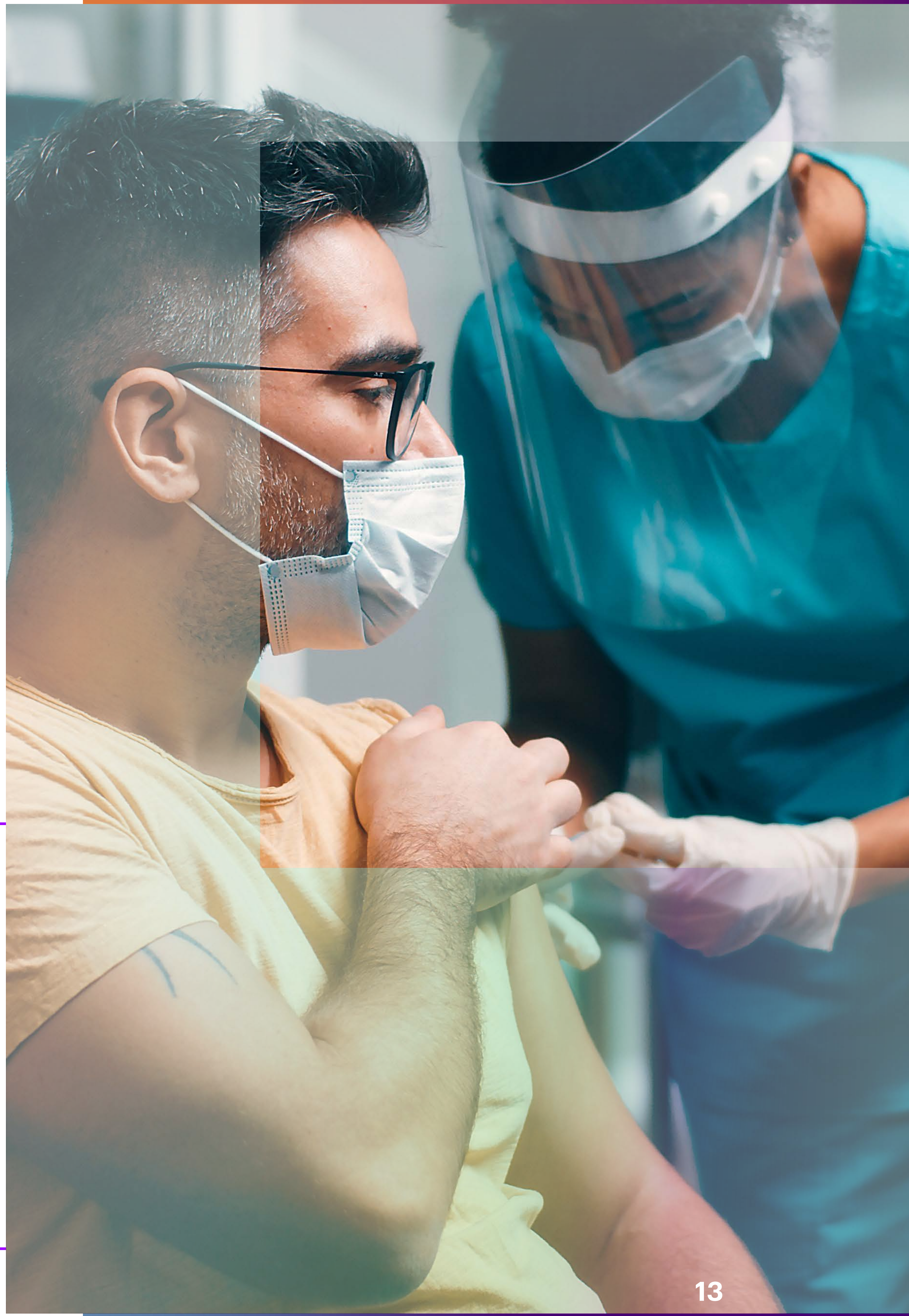
In public health surveillance, IA could significantly improve infectious disease reporting by automating the collection and transfer of relevant health information from EHRs. When a health worker records a particular symptom or disease case in a patient's EHR, the IA system could automatically send the data directly to CDC or other agencies, eliminating the administrative burden currently required for reporting. IA systems could also scan and interpret lab reports or clinical notes to uncover disease cases that might otherwise elude health officials and trigger reports to state and local authorities.

IA not only automates predefined, repeated tasks, but also allows the system to learn and adapt. Powered by artificial intelligence and machine learning, an IA system for extracting data from unstructured text can go beyond simple optical character recognition, leveraging NLP to understand context, reduce noise, and improve accuracy.

By employing IA solutions, public health agencies can produce more complete and accurate assessments of disease burden and trends while simultaneously enhancing operational efficiency – eliminating manual, repetitive work and allowing human workers to focus on higher-value tasks.

Applying government AI standards to intelligent automation solutions

IA solutions that rely on artificial intelligence must follow federal guidance for the design, development, testing, and continuous monitoring of responsible and trustworthy AI, including the Government Accounting Office's Accountability Framework for Federal Agencies, the White House Blueprint for an AI Bill of Rights, and the National Institute of Standards and Technology's AI Risk Management Framework.



[21] Workplace Perceptions and Experiences Related to COVID-19 Response Efforts Among Public Health Workers (cdc.gov)

An action plan for alignment and governance

As federal agencies define and implement a public health surveillance system that integrates rich, interoperable data to power robust analytical tools and IA solutions at scale, long-term success will hinge on alignment with key data partners and clear governance.

By including state and local agencies, HIEs, data aggregators, laboratories, and/or other data partners and focusing on discrete use cases, federal public health leaders can pursue an iterative approach to defining and testing solutions – while simultaneously supporting effective change management across public health stakeholders.

Initial steps

1. **Define one or more discrete, priority use cases** – such as HIV, cancer, diabetes, or foodborne illness – to test and demonstrate the potential value of data solutions.

2. **Select data partners** for the initial use cases whose data sources can be integrated into a data mesh solution or migrated to a cloud architecture so they can be accessed.

3. **Create a robust, participatory governance framework**, including a strategic framework to address policy, technical, and operational considerations, as well as data governance to ensure HIPAA compliance for patient health data and align standards for non-health data sources.

A vision for the future of public health surveillance

As public health agencies integrate – and act on – lessons from the COVID-19 pandemic, strengthening America's surveillance system represents the highest priority.

By investing in next-generation infrastructure and expanding the universe of available and interoperable data, agencies can establish an analytical pipeline with unprecedented robustness. This pipeline would fuel models and simulations with sufficient power to derive real-time insights – for better policy and programs focused on prevention, control, and response. Armed with the power of intelligent automation, public health agencies can implement these advances without further taxing the workforce – effectively doing more with less.

These strategic investments hold the key to real-time surveillance data and insights that allow our leaders to understand disease burden, predict future risk, develop and evaluate prevention and control strategies, and – ultimately – save lives.

Authors

Kenyon Crowley
Managing Director,
Health Analytics

Accenture Federal Services

Kasey Morris
Manager, Health Analytics

Accenture Federal Services

About Accenture

Accenture is a global professional services company with leading capabilities in digital, cloud and security. Combining unmatched experience and specialized skills across more than 40 industries, we offer Strategy and Consulting, Technology and Operations services and Accenture Song — all powered by the world's largest network of Advanced Technology and Intelligent Operations centers. Our 738,000 people deliver on the promise of technology and human ingenuity every day, serving clients in more than 120 countries. We embrace the power of change to create value and shared success for our clients, people, shareholders, partners and communities. Visit us at **accenture.com**.

About Accenture Federal Services

Accenture Federal Services is a US company headquartered in Arlington, Virginia, and a subsidiary of Accenture LLP. We empower the federal government to accelerate change. Together we help clients achieve greater outcomes that make the country stronger and life better for all people, and advance US leadership in the world.

We are privileged to serve every Cabinet-level department and federal organizations serving mission-critical priorities in national security, defense, safety, public and military health, and civilian sectors.

Our 13,000 people draw out the best of Accenture's global network in nearly every technology and industry to bring even more power and proven commercial innovation to our mission-tested solutions built with advanced R&D, disruptive technologies, data, and human-centered design. So our clients can create more value today and forge a bolder future for everyone. See how we make change that matters at at **accenturefederal.com**.