Modernizing Data and Analytics at the NYC Health Department

Center for Population Health Data Science
Two-Year Lookback Report
2023-2025

NYC Health

Introduction

Mamta Parakh
Chief Population Health and Data Officer, Director

Gretchen Van Wye
Chief Epidemiologist, Deputy Director

Data is our superpower in public health. It helps sound alarms, spur action, and drive planning, programs, and policy for health. The NYC Health Department, the nation's largest and oldest municipal public health agency, has a 220-year history of understanding local information at a deep level to inform action and address public health challenges of the time.

Since our founding, health issues have become more complex, and we see differences in health outcomes between groups that reflect social inequalities. Gaps in data and between data systems can keep us from seeing the full picture of health in neighborhoods and groups. Today, we are charged with bringing together data from many sources to create a more complete and timelier picture, so we can direct help where it is needed most.

The Center for Population Health Data Science (CPHDS) is the hub of the Health Department's data modernization work. We are building connected, resilient, adaptable, and sustainable response-ready systems that will help us do our everyday work more easily and be better prepared to respond to future public health emergencies.

We look forward to the future for data in NYC, as we appreciate the journey to get us here. With gratitude to our dedicated colleagues in CPHDS, and across the Health Department, we share this reflection on our first two years.

Center for Population Health Data Science

In CPHDS's role leading the Health Department's data modernization efforts, our mission is to **unite people, services, and technology to transform data into a trusted and shared resource** that strengthens public health practice and maximizes impact.

More than 350 people work together across five bureaus to carry out this mission. Three bureaus — Epidemiology Services, Vital Statistics, and the World Trade Center Health Registry — conduct essential data collection, health monitoring, and service delivery that have anchored public health for decades — and longer. Our Administration bureau ensures operational excellence. Our Data Technology and Strategy bureau builds modern data infrastructure that integrates legacy systems, creates unified representations of data, embeds equity into the data structure, and enables responsible applications of AI and advanced analytics.

In CPHDS, epidemiology meets data engineering, user experience design intersects with governance, and public health practice drives technical innovation. Together, we aim to achieve our vision to create a **continuously learning public health system** that harnesses data and technology to deliver responsive services, advance health for all, and drive equitable outcomes.

Our founding strengths

Deep epidemiological expertise

Teams of epidemiologists, biostatisticians, data analysts and public health professionals brought decades of experience assessing health and translating data into informed solutions

Trusted data operations

Well-established systems for collecting, processing, and analyzing vital records, health registries, and survey data, supported by strong research partnerships and a history of sharing

Solid public health foundation

Strong, time-tested capabilities in data collection, analysis, and interpretation, grounded in rigorous statistical methods and sound data management practices

Challenges we faced

Technical fragmentation

Decades of program-specific built data systems lacked interoperability, forcing researchers to spend more time locating and preparing data than analyzing it

Capacity for modern methods

Limited expertise and infrastructure for modern data engineering, cloud technology, and enterprise data management

Organizational dynamics

Distinct workflows and tools across divisions made it difficult to coordinate and integrate shared data capabilities across the Health Department

Early Investments in People and Systems

Building on CDC's Data Modernization Initiative key priorities and Public Health Data Strategy, we identified five interconnected priority areas to organize and focus our work. For each priority area, we articulated a vision statement describing where we aim to be within five years and made strategic investments to begin moving in that direction. This framework guides decision-making while remaining adaptable. Our agile approach allows us to test solutions, learn from what works, and adjust course while staying anchored to these core priorities and vision statements.

Our approach champions data modernization as a collaborative effort across the agency, with CPHDS serving as the hub. We actively seek input through cross-divisional working groups, bureau and program-focused discovery sessions, and regular engagement with data stewards, program partners, and division leaders. Our work is grounded in ongoing dialog about stakeholders' needs, priorities, and challenges.

Five-Year Vision for Data

Data for Equity & Social Justice

Data is used to shape fair policies and systems, empower New Yorkers to advocate for health, and ensure that public health decisions are focus on equity.

Data Infrastructure Modernization

A secure, adaptable, and cutting-edge data infrastructure enables seamless access for analysis, trusted sharing, and world-class security.

Public Health Workforce Development

A highly skilled, collaborative, and mission-driven data workforce is cherished, where epidemiologists, data scientists, engineers, and public health practitioners work together to turn data into impact.

Advanced Epidemiology & Analytics

Real-time data and advanced analytics are leveraged to guide policy, prevent disease, and drive smarter, more effective public health action for the city.

Strategy & Partnerships for Modernization

A connected data ecosystem enables partners to share insights, researchers to access integrated data, and real-time information to fuel a healthier city.

Data for Equity and Social Justice

What We've Built and What It Means

What's now possible

A hub for better data, tools, and guidance to understand health outcome differences across communities, identify where to focus resources, and improve how we collect and use information to serve all New Yorkers fairly — with more development underway.

What it means

Equity advances when we move from aspirations to actions to impact. Collection and analysis of data in a way that is equitable enables us to document inequities so that we can address them. Race/ethnicity guidance will help divisions apply evolving national standards consistently, with common definitions, across diverse program contexts. Neighborhood prioritization methodology focuses investments on communities with compounding disadvantage. Wealth gap insights shed light on why families face different health trajectories.

As we complete these guidelines and methods, we can begin embedding them into our data infrastructure, moving toward equity-centered analysis as the default.

Select initiatives

- **Wealth gap survey** assessing assets and debts among NYC to understand relationships between health and wealth across racial/ethnic groups.
- **Immigrant Health Report** describing health challenges that immigrants face and specific interventions needed.
- **Neighborhood identification and investment methodology** based on Healthy NYC priorities with consideration for historical structural harms.
- **Impact of new Census race/ethnicity categories** on our data and guidance to support implementing them.
- **Equity skills survey** (second agency-wide assessment) found increased skills in 2024 compared to 2019 and developed recommendations to strengthen capacity.
- **Excellence in customer service** while receiving and completing a record number of birth certificate and vital record amendment requests.

Data Infrastructure Modernization

What We've Built and What It Means

What's now possible

From new online data resources that make information easier to find and use, to the development of efficient and secure systems behind the scenes, we are creating the infrastructure that allows trusted, high-quality data to drive timely insights and smarter public health action.

What it means

The Enterprise Data Platform (EDP) is built to tackle specific pain points in integrated data and analytics workflows across the agency, for day-to-day work and during emergencies. With security approvals and operational infrastructure, new use cases and data sets will be onboarded so data pipelines can be reused — not rebuilt from scratch — with each new project, reducing time to insight. With data integrated from across systems, we can interpret information at the community, block, and address levels and localize public health action for greater effectiveness.

Select initiatives

- **Cross-functional technical capacity through OneTeam collaboration** among CPHDS data engineers, IT infrastructure specialists, and Microsoft cloud experts. This partnership produced key advances: The agency's first scalable system for managing data access across divisions and sensitivity levels, engineering best practices using development pipelines with peer code review that enable quality control and knowledge sharing, and increased team fluency in modern cloud technology stack.
- **New public-facing data products:**
  - Provisional Birth and Death Data allowing for more timely estimates with faster ongoing updates.
  - Respiratory Illness Data page combining COVID-19, RSV, and influenza data to improve findability and interpretability, align with reporting norms, and reduce ad hoc reporting and data requests.
- **Digital mother/parent worksheet pilot for parents** to provide information for newborns.

Public Health Workforce Development

What We've Built and What It Means

What's now possible

We now have a clear, shared mission, routinized communication structures, stronger teamwork and connection across bureaus, and new technical and data skills. We are connecting public health and data science, sharing knowledge, and working together more effectively.

Building shared identity and understanding across diverse fields takes intentionality, investment, and time. Our teams are creating a new shared language and adapting to new ways of working together.

What it means

We have built a small, dedicated data technology team that understands public health and modern data practices. We've created a pathway for epidemiologists and public health researchers to learn, leverage, and benefit from modern technical skills. With a focus on expanding capacity across the agency, we are poised to support multiple concurrent data modernization use cases.

Select initiatives

- **New mission and vision statements** developed through a series of collaborative all-staff ideation sessions allow everyone to see themselves in the transformation
- **Deeper cross-bureau awareness and understanding** through regular all-staff meetings, informal community gatherings, onboarding and training plans, and routinized communications
- **Bridging the gap between data science and public health** through dialogue, presentations, and trainings (for example, public health 101, causal inference, Databricks and DBT, and knowledge graphs).
- **New personnel with skills** in data engineering, data governance, product management, and user experience research.
- **Representation on key internal technology governance bodies** to align data modernization with enterprise technology strategy.

Advanced Epidemiology and Analytics

What We've Built and What It Means

What's now possible

We are building the analytical capabilities to monitor health in real time, predict trends, and evaluate intervention impacts — connecting newer methods with public health practice.

What it means

Today, many epidemiological analyses still rely on single data sources and descriptive methods. The new infrastructure positions us to move up the hierarchy of study designs — linking datasets for case-control and cohort studies, applying predictive analytics to anticipate outbreaks, and building prescriptive models that simulate intervention impacts before implementation. This requires systematically aligning analytic agendas with infrastructure build and training. Our proof-of-concept projects show what is possible. Scaling these approaches across the breadth of our epidemiological work — while maintaining excellence in core public health — is the substantial work ahead.

Select initiatives

- **Population Health Impacts Surveillance System** to monitor impacts of major policy shifts and societal changes.
- **Quality and utility of EHR data for health monitoring** and analytics with the local health information exchanges.
- **Real-world patterns of opioid use disorder treatment** across multiple linked datasets under a new NIH grant.
- **Key health indicator tracking through Community Health Survey** with nearly 10,000 respondents.
- **Agent-based simulation models** with academic researchers to estimate life expectancy gains of HealthyNYC investments.
- **New predictive modeling tools** for health impact of policy changes and early warning systems.
- **Custom vision and language models** to process unstructured data across several public health use cases.

Strategy and Partnerships for Modernization

What We've Built and What It Means

What's now possible

We now have stronger partnerships and expert guidance that help us share knowledge, safely manage sensitive data, and work with other agencies to connect health and social services for New Yorkers.

What it means

These relationships are building blocks for the connected data ecosystem we're creating and help us stay in touch with the communities and systems we serve. External partners and peers bring perspectives that sharpen our strategy and help build a collective voice to advance data modernization. We've planted seeds that will grow into productive, lasting collaborations as specific needs emerge.

Select initiatives

- **Minimum necessary guidelines** to ensure that when New Yorkers' data is shared it remains safe and protected.
- **Relationships with external partners:** NYC agencies and the Office of the Mayor (NYC Health + Hospitals, Office of Technology Infrastructure, Center for Innovation Data Intelligence), New York State Department of Health and other state-focused DMI groups (NYeC), and the CDC Office of Public Health Data, Surveillance, and Technology.
- **Data Transformation Advisory Committee**, a diverse group of external data leaders and technology experts, who advise on the strategic direction of data initiatives.
- **Sustainable, long-term value exchange models** and ways to bring together public health expertise and infrastructure with private sector resources to achieve health outcomes.
- **Support of Social Care Networks to implement the 1115 Medicaid Waiver** where health-related social needs referrals are covered by Medicaid to integrate social determinants of health into health care.

Lessons Learned and Looking Forward

Foundation building takes longer than you think...

...and that's okay. Anything worthwhile does.

No heroes

Diverse roles and perspectives are necessary. Encourage speaking up and build safety in dissenting and thinking outside the box.

Bilingualism

People who can speak R/Python and public health/social determinants of health are key. Data people need context, epidemiologists need pipelines.

Relationships drive sustainability

Workshops, trainings, offsites, and conversations are a key part of development.

The Road Ahead

With the foundation for transformative infrastructure in place, strategic investment in data modernization will create new opportunities to better serve New Yorkers. Data modernization is an activity of continuous improvement, not a discrete task to complete. The capabilities we build today open possibilities for tomorrow that didn't exist before. By organizing our data well and building in the right protections across our priority areas, we build readiness for responsible AI and use of new technologies safely and effectively to advance public health.

Continued investment will give city leaders clear, useful data to make informed decisions about priorities and resources, helping us achieve our public health goals. Consistent ways to measure results across programs will help us show communities the real impact of our work. New Yorkers will experience smoother, more connected interactions with public health services, reducing confusion and building trust. These improvements will allow us to work with the efficiency and openness that the public expects and deserves.