DMI

Data Modernization INITIATIVE

Stories from the field

Enterprise approach to data systems modernization

Innovative uses of data

Communication strategies

SCALING UP DATA STRATEGY AND TECHNOLOGY IN THE BIG APPLE

The New York City Department of Health and Mental Hygiene leverages its history of technology innovations for agencywide data strategies

New York City (NYC) has an ethnically, linguistically and economically diverse population of more than 8.3 million people. As the largest and most densely populated city in the U.S., disease can spread quickly. Other public health challenges include obesity, diabetes, heart disease, tobacco addiction and substance use, as well as the threat of bioterrorism. To address these challenges equitably and protect the health of New Yorkers, the NYC Department of Health and Mental Hygiene (NYC Health Department) has a history of innovating and evolving its public health data systems and practices.

Technological advances for data systems

Twenty years ago, the NYC Health Department built a Master Child Index (MCI), combining its immunization and blood lead registries. The Citywide Immunization Registry is seeded with birth records from the NYC Health Department's Bureau of Vital Statistics and receives immunization records from healthcare providers in NYC. The blood lead registry receives laboratory and provider reports for blood lead levels in children. These data systems were not always easily linked, so the technology team built an artificial intelligence (AI) model to conduct probabilistic identity matching—a novel approach in 2004.

The NYC Health Department data modernization team designs an innovative, cloud native, central data infrastructure.
Photo credit: Henry Kueppers

Angel Aponte, Director of Information Strategy for the NYC Health Department's Division of Disease Control says, "The rationale for building the Master Child Index was to … integrate the data, to provide the healthcare providers with a single view of that child's health in the agency's online registry web application."

The agency built the system "from scratch, using available technologies at the time," says Aponte. In contrast, today's technological advances in free, open-source software provide easier access to better AI and machine learning tools, so Aponte's team no longer needs to build all the foundational technologies. Because of these advances and lessons learned from developing and refining the MCI, the NYC Health Department is using these new technologies to implement more sustainable, next-generation matching software.


Angel Aponte

Aponte says, "We want to grow the index to be more than just children. We would like to integrate data for all of NYC's health systems, to help inform how to best support the health of our population. We will have to prioritize what systems we want to integrate and probably take them on one at a time, add another one, then add another one until we have a larger view of a person's health here in New York City."

"Due diligence and processes to secure the data take time," says Aponte. "The data that we collect are very confidential. We're the stewards of that information, and we need to make sure that we protect it for New Yorkers, the agency and [the] city."

Aponte says they should also implement "data lake house" architecture for pooling data across sources. It would be a flexible technology for integrating data from multiple different sources to support analysis about how to best improve the health of NYC's population. The NYC Health Department's Division of Information Technology did a proof of concept for a lake house using Microsoft Azure, and Aponte says they are looking to expand on that.

> "A data lake gives you a lot of flexibility for receiving and ingesting data from structured, semi-structured and unstructured sources. A warehouse gives you the ability to slice, dice and analyze data from all of those different sources in a very flexible way. When you add the flexibility of data ingestion and the flexibility of data analysis, you build a data lake house," says Aponte.

The Division of Disease Control has also been evolving their electronic disease reporting infrastructure. During the height of the COVID-19 pandemic, the NYC Health Department found this system needed improvements. The system receives electronic lab reporting data as well as "one-record-at-a-time reports through a web application … for disease surveillance in particular," says Aponte. The system helps clean and standardize codes and route data to the appropriate destination.

Aponte notes disease surveillance can include all sorts of different data. "So, we've got communicable disease, we've got HIV, hepatitis C and sexually transmitted infections, tuberculosis, [and] we have vaccine-preventable diseases. [They are] all using different surveillance systems, and the data is somewhat linked between the surveillance systems, but we can do better."

Aponte says his team plans to work with the Division of Information Technology and the Center for Population Health Data Science to integrate the new disease reporting infrastructure with the person index the agency is building. That way, "We'll have all of that surveillance information from all sources, and we're going to add electronic case reporting [and] health information exchange data as well. We'll have a single fully consolidated view of the health of our population." Aponte says the project timeline for the new system just started and anticipates it will take three years.

Aponte notes, "Our job is to protect the population from health risks, innovate, and implement new and better infrastructure. You can move quickly and make mistakes in a development environment with synthetic data, but if you move too quickly with real patient records, you pay a hefty reputational price for mistakes. We have to get results and not put the information that we protect at risk."

Data sharing and matching challenges

Despite being ahead of the curve for data systems, the NYC Health Department still experiences data challenges other jurisdictions face. Charles (Chip) Ko, Senior Director of Integrated Surveillance and Data in the Bureau of Epidemiology Services, gives the following example: During the COVID-19 pandemic, the NYC Health Department used complex processes to match different datasets—they matched COVID-19 data to data from vital statistics, the immunization registry and other syndromic surveillance programs. According to Ko, this work was vital during the height of the COVID-19 pandemic and "really allowed us to better track and describe the outcomes associated with COVID infection."

Ko adds, "The problem, though, is that this required so much staff time, conducting all these one-off matches, monitoring those processes and addressing quality assurance issues. The hope with data modernization is that we would be able to access that data much more quickly ... [so we] can spend more time on in-depth analyses where we can gain insight much more quickly."

Charles (Chip) Ko

"The hope with data modernization is that we would be able to access that data much more quickly ... [so we] can spend more time on in-depth analyses where we can insight much more quickly."

Data Modernization Initiative (DMI) Unit

The Data Modernization Initiative (DMI) Unit was created in 2022 to address these data issues and "strive toward connected, adaptable response-ready systems that use public health data to drive action and equity," says Ko.

The DMI Unit is part of the NYC Health Department's newly launched Center for Population Health Data Science, a center aimed to catalyze the agency's role as the "city's health strategist" with a goal to link public health, healthcare and social service data. The DMI Unit currently includes over 20 employees, with subunits in data governance, data acquisition, and data science and engineering. However, Ko stresses that not all the work happens within the DMI Unit. "It is really a cross-agency collaboration," he says.

Collaborative forums

One of the DMI Unit's first initiatives was to create a data modernization advisory group, with representation across the NYC Health Department. The purpose of the group is to ensure "we have representatives, cross-agency, to ensure that everyone has a voice at the table," says Ko.

The advisory group meets monthly and includes peer-to-peer conversations to expand understanding of different data workflows or methods across divisions. Different team members share their work, discuss challenges and successes, and answer questions. Past discussions have included record matching and a deep dive on geographic information system (GIS) analyses.

To form the advisory group, Ko's team asked divisional leadership to nominate one to two people from their divisions. "One person that may have a day-to-day analytical role. And then another person who might be a strategic thinker. We thought those two would complement each other and be a good pairing for this advisory group," says Ko.

Team members meet regularly to discuss DMI strategies and projects. Photo credit: Henry Kueppers

The advisory group holds informational sessions led by DMI members on topics such as data governance, data trust and data maturity assessments. Ko says the advisory group acts like a conduit for delivering their DMI updates to the other respective divisions across the NYC Health Department.

To ensure the DMI Unit was communicating even more broadly across the agency, the team also set up monthly roundtables for sharing key updates with executive leadership. "That includes our Commissioner of Health, Chief of Staff, Chief Operating Officer and leadership from other divisions who aren't normally in our regularly scheduled calls. We have a space to raise any concerns we had or where we might need support from the executive leadership team,"says Ko.

Linking technical advances across the agency

Ko reflects on how his team seeks to understand technology innovations agency-wide: "How do we possibly know everything that's going on regarding data modernization? … We've set up communities of practice to foster these organic collaborations and knowledge transfer across the divisions."

The DMI Unit also hosts data science and data engineering office hours, which are attended by around 40 people every other week. The format alternates between technical presentations and technical support with the purpose of advancing data science capacity across the NYC Health Department. Ko says, "People drop in and come with questions or pain points they're trying to address … everyone just comes together and thinks through different solutions or asks 'Have you tried this? Have you tried that?' It's a good opportunity for sharing ideas and methods."

Ko also says it has been important to find "collaborative champions" across the NYC Health Department, such as Aponte. Ko says Aponte's experience at the agency and his work on the next-generation matching software project has been valuable for the DMI Unit. "He has much experience at the agency and is working on [projects] that we really want to gain insight on. We're working with him, learning from his team on how we can implement something similar agencywide," says Ko.

Aponte says his goal is to "build a pipeline of innovative projects and useful infrastructure projects within one division and then hand them off for greater agency use." He adds, "If I do a good job of building infrastructure for disease control, they should be able to scale to things that are larger than just disease control in New York City, especially since our population is so large."

Aponte also notes the DMI Unit's work on governance structures will be critical for integrating information—getting a holistic view of a person's health—while simultaneously protecting secure data. He says this

work is challenging: “It’s policy and process. It’s not something that a technologist is necessarily trained to do … it’s something that requires a different skill set. I’m fortunate to have partners who are doing this work.”

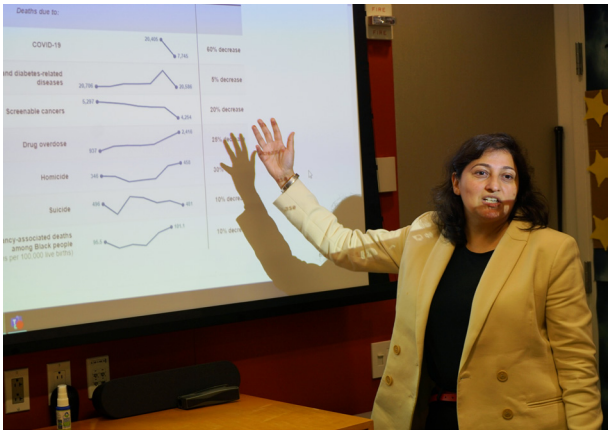
Newsletter updates

To keep staff and leadership updated, the DMI Unit sends a regular monthly newsletter to the executive leadership team, the advisory group and any staff across the NYC Health Department interested in learning about DMI. The newsletter’s purpose is to highlight data modernization projects and relevant policy updates while increasing agency awareness.

Ana Gutierrez

Ana Gutierrez, a Strategic Initiatives Manager in the DMI Unit, creates the newsletter. She develops the content and notes it is “a grassroots effort of folks that read it.” She gives the example of the public health lab sending information about an innovative project. Gutierrez says, “That kind of information may not be within our purview, but it also advances data modernization.”

The newsletter is formatted so the activity updates are aligned with the Centers for Disease Control and Prevention’s (CDC) DMI priorities. “By setting that as the [standards] and then adding updates within it, we’re making connections for recipients to know why we’re doing certain activities. For example, we’re conducting a data maturity assessment, and it aligns with priority one. It helps communicate data modernization in a way that’s not intimidating, so there isn’t much of a roadblock to entering the conversation,” says Gutierrez .



Mamta Parakh, Chief Population Health and Data Officer, presents on NYC’s vision on how to improve life expectancy and create a healthier city.
Photo credit: Henry Kueppers

Gutierrez says the language they use may emphasize how a DMI project can help reduce the pain points that public health staff face in their work: “Connecting a project to the pain points that people are familiar with allows it to be more accessible and in less technical language.”

DMI projects

In 2022, the DMI Unit’s advisory group identified five main priorities using the CDC’s self-assessment. These priorities are increasing capacity for data linkages and data standardization, coordinating governance for all data assets at the agency, improving data sharing internally and externally, increasing workforce capacity, and improving technology infrastructure.

Currently, the DMI Unit’s focus is planning a blueprint for a central data infrastructure. The design of this infrastructure will be informed by two assessments being led by the team: a data maturity and a data governance assessment. The governance assessment would be used to understand the NYC Health Department’s readiness for an enterprise-wide data catalog while also gathering insights into current governance practices at the agency with the goal

Promising practices

- Develop an agencywide definition of data modernization to help create shared language and understanding.
- Create communities of practice to foster organic collaboration and knowledge sharing.
- Send regular updates such as a monthly newsletter. To appeal to different audiences, connect data modernization projects as solutions to pain points.
- Explore free software; technology itself does not have to be a big investment.
- Build technology solutions with scale and an enterprise approach in mind.

of moving toward a centralized data governance framework.

The catalog would include "all data assets that we have at the agency." Ko says the system would also ideally help them manage data use agreements: "Having them all in one place instead of having them siloed within the different program areas."

The immense work for a new central data infrastructure includes building internal capacity, investing in a shared cloud infrastructure and building out core data integration pipelines.

Obstacles in data modernization

On the technology side, Aponte reflects, "Everybody wants to see results, and they want to get to the end goal, and you've got this beautiful kind of presentation at the end of it. But there is a significant amount of work and planning that needs to happen to get to that end result."

Technical innovations are also not always visible. Aponte acknowledges that advancements in technology that support public health are not always visible to the public: "It's known in the agency, but the public would not necessarily know. Many people may not even know that an immunization registry exists except that we have a web application called My Vaccine Record, where a person can look up their own or minor child's immunization record to show, 'Yeah, I was vaccinated for COVID.' They don't see all the other infrastructure that supports health care provider vaccine orders, accounts for vaccine use and provides that immunization information to the electronic health records when they see a different doctor."

NYC Health Department defines data modernization as any NYC Health system, project or person enhancing efficiency and interoperability of data and data systems to support use of public health data to inform policy and interventions for the health of all New Yorkers.

Aponte said this is historically a challenge for public health: "Public health does its best work when our population is well protected—we're often not noticed when things go well, and we tend to get noticed when there's a problem."

To retain technology positions, Aponte advises health departments to hire young technologists who can learn on the job and grow with an agency. He says an advantage for NYC is that technologists are integrated in the public health department rather than being centralized in an external agency: "A lot of the health department's work is technology…you have to have those skills in the agency to really move the needle."

Success factors

To make progress in data modernization, Ko says communication is critical: "We found it's really important to use shared language because data modernization is so broad, and it entails so much. It means something different to everyone." The NYC Health Department has its own data modernization definition that Ko says is helpful for project focus and describing DMI work to staff across the agency. Ko adds, "Having shared language is also really important in this work because we're speaking with technical and nontechnical audiences. Trying to find that balance [between technical and nontechnical language] is also important." Laney says obstacles can be delays in data sharing agreements, funding, and staff bandwidth. "The staff has set job requirements, and then the modernization adds more to the daily workload."

Another factor for success is collaboration. Ko says it is important to create spaces where collaboration can occur and elevate spaces like the community of practices. He also says it is beneficial to identify collaborative champions at the NYC Health Department and learn from their work: "Angel Aponte is a huge collaborative champion here and we've learned so much from his team."

Ko mentions being agile and able to adapt to the needs of the organization is important because data modernization is not going to look the same at every health department.

Advice for other jurisdictions

Aponte says free software is the underlying foundation for everything they have built. He suggests exploring free software such as Python, R and the Linux operating system. He adds: "There are free database technologies as well, free web application technologies from Apache, Tomcat and PostgreSQL. We build public health infrastructure on the foundation of free software."

Aponte says public health work is interdisciplinary and problems do not necessarily have an elegant solution: "You often have to make the best decision you can with incomplete information, [and] you may not have all the answers right away. But what you need to do is move the needle and improve incrementally. If you keep making incremental progress, you will get to the goal eventually. Every change that you make doesn't have to be revolutionary."

Gutierrez agrees, saying the NYC Health Department "has big initiatives that are trying to redesign a lot of the ways we do things, but then we also have smaller conversations that provide small improvements in day-to-day work."

Ko says starting with an assessment is a great step toward data modernization. He refers to the NYC Health Department's data maturity assessment: "We really feel that this assessment is going to help us prioritize the projects and the next steps for our DMI roadmap. Maybe that's a tool that other health departments can use to understand what to tackle first." He references the Public Health Informatics Institute's informatics-Savvy Health Department self-assessment tool as a place to begin.

Looking toward the future

Ko says, "What we want from data modernization is to gain insights about our city's population and their health needs—to use those insights to plan and help the health of all New Yorkers."

"Data modernization is a long-term investment in the public health technology infrastructure. Public health practitioners collect, integrate and distill high-quality, actionable information from various sources to inform program staff and leaders at our health department, in our city, [in] our state and in our country so that they can make the best decisions possible to support our population. We use data to inform public health interventions that improve the health of our population. I think that modernizing our infrastructure will help executives make the best decisions, so our agency can better protect our population," says Aponte.

To carry on the work of data modernization, Aponte says a sustained level of resources is needed. "We need to commit to providing long-term sustained resources for public health and for our workforce. I've seen a cycle in my time: You've got an emergency and a large influx of resources, and then the emergency ends, we get things under control, and the resources fall off a cliff and dwindle down until you get back to the baseline that you started and you often end up even lower than the baseline. We need a sustained commitment. It's not pandemic-level resources. It's not pre-pandemic-level resources. It's somewhere in the middle, where we can ensure that we're as modern and well-prepared as we can be."

Additional resources

- NYC Health Department website. https://www.nyc.gov/site/doh/
- Public Health Informatics Institute Informatics-Savvy Health Department Toolkit. https://phii.org/course/informatics-savvy

Key takeaways

- Twenty years ago, the Master Child Index (MCI) was built using innovative AI technologies at the time.
- With technology advances and lessons learned from the MCI, the NYC Health Department is using new technologies to implement next-generation matching software and a future central data infrastructure.
- The DMI Unit was created to strive toward connected systems that use public health data to drive action and equity.
- Groups meet regularly to share work, challenges and successes. A monthly newsletter also builds awareness; highlighting DMI news and relevant policy updates.
- Innovative data projects are created with the intention of scaling them up and making them applicable across the entire NYC Health enterprise.