Top 5 lessons from New York City's digital transformation journey

By Dr Amen Ra Mashariki

As the Global Director for the Data Lab at the World Resources Institute, and the former Chief Analytics Officer for the City of New York, Dr Amen Ra Mashariki is one of the most experienced data strategists in the world when it comes to leveraging urban analytics and data science to solve complex problems.

——

In this series of essays, Dr Mashariki reveals the top 5 lessons from New York City's digital transformation journey and shares best practice data strategies for creating smarter cities.

Contents

How data failed during September 11 and other large-scale data-sharing crises

I was recently asked a provocative question by a student. "What happens when data fails?"

I took a second to gather my thoughts in order to give an answer that would both challenge and engage. My mind was racing during the four seconds it took for me to say: "That is a very good question."

I raced through a number of possible responses in my mind and thought about answering his question from a technical perspective, which is to discuss what happens when data systems fail or crash.

I thought about answering his question from a human perspective, which is to discuss what happens when the data solutions we deploy don't work to help the intended audience. I thought about Cambridge Analytica and Facebook and how data privacy violations and the misuse of data can be seen as 'data failing'.

At the end of my four seconds, my answer hit me like a ton of bricks and I said: "I saw what happens when data fails during my tenure at New York City during 9/11."

What I discussed then was the fact that the inability to share data with those who need it at all times caused challenges during both manmade and natural catastrophes.
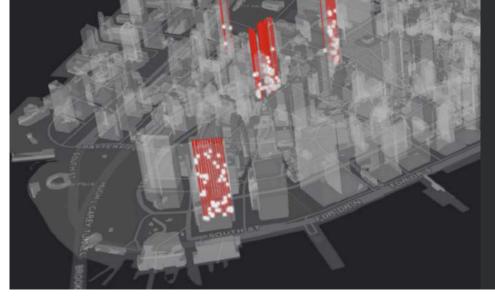
I used New York City as an example because I worked on issues like these during my tenure at the city. And, as a native New Yorker with family still living there, these issues are very important to me.

The lack of data sharing leading up to 9/11 is outlined in the 9/11 commission report:

"The U.S. government has access to a vast amount of information. But it has a weak system for processing and using what it has."

The system of 'need to know' should be replaced by a system of 'need to share.'

There are many anecdotal reports on the ground of emergency responders struggling to get the data that they needed to do their jobs quickly and effectively.

In order to help New York City prepare for a crisis, Dr Mashariki ran a data drill. In the hypothetical scenario, the city lost power for 14 days. The team mapped all of the working elevators in the city to answer the question: "How many elevators are down due to the blackout and can we quickly route emergency vehicles to the required locations?"
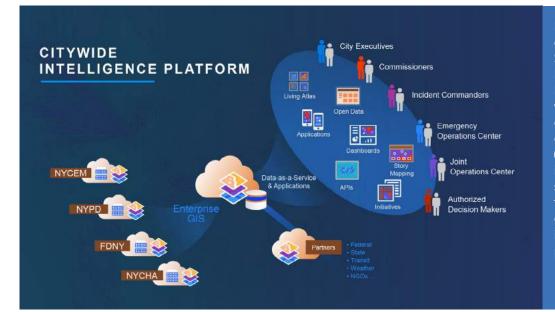
Hurricane Sandy

Data also failed during and after 2012's Hurricane Sandy. The Sandy After Action Report explicitly stated that "the city should significantly improve its collection and synthesis of data on the provision of essential services throughout the city, including power, gas, and telecommunications."

In 2013, The United States District Court Southern District of New York found the city liable for failing to provide meaningful access to people with disabilities to its emergency preparedness programs and services. As a result of that decision, the city was required to create a Post-Emergency Canvassing Operation (PECO) plan to rapidly survey households after a disaster to assess and identify the critical needs of people with disabilities.

During a canvassing operation to survey the households, canvassers were required to go door-to-door carrying a mobile survey tool to input resource requests and refer those requests to appropriate partners for resolution. Resource requests included but were not limited to: food, water, electricity, medical care and durable medical equipment.

But the city did not have a readily available database of people with disabilities.

The problem wasn't that the data didn't exist; the problem was that data was spread across multiple agencies in disparate databases with no chance of it ever being shared or smartly integrated for use during an emergency.

In the aftermath of Hurricane Sandy, Dr Amen Ra Mashariki worked to connect all of the GIS (geographic information systems) to create a city-wide intelligence platform.

New York City's legionnaires outbreak

Data also failed during the summer of 2015 when a terrible outbreak of Legionnaires disease in New York caused 12 deaths and infected 112 people.

Ultimately, city health inspectors and first responders found that the bacteria was spreading via cooling towers. This was difficult to discern at the time since the city did not have a database of cooling tower locations.

As this example demonstrates, it's the 'unknown unknowns' that will hurt you.

The examples that I mentioned above have either been rectified or are being worked on.

I think these examples are relatable to many of us because these scenarios of 'data failing' New Yorkers is something that happens across every government agency, everywhere.

So what are we doing about it?

I am seeing governments hiring Chief Data Officers and data scientists so the right people are at the helm to ensure that data doesn't fail again in serious situations.

I have the pleasure of knowing and working with many data professionals in government and I am comfortable knowing that federal, state and local government – in the United States, Australia and all around the world – are headed in the right direction to make sure that data doesn't fail us in a catastrophic way ever again.

LESSON 2

Engage citizens with open and transparent public data portals that offer insights into the operations of government

How open data and location-based analytics saved residents millions in parking tickets

During my role as Chief Analytics Officer for the city of New York, the Mayor's Office of Data Analytics (MODA), along with the Department of Information Technology and Telecommunication (DoITT), created the 'NYC Open Data Team.'

These departments created lots of strong open data tools connecting New Yorkers and NYC government in many ways and on many different platforms.

This essay is about one of those instances. It is the telling from my perspective of a big open data story that broke in NYC.

Ben Wellington is the creator of a blog called IQuantNY, as well as a data scientist at a hedge fund in New York. Through his blog, Ben uses NYC Open Data to do some very insightful, and sometimes humorous, analysis about the city.

One afternoon, Ben emailed me while I was at work. He gave me a head's up about a blog he had written focusing on an erroneous action by the NYC government.

> He had used Open Data to uncover the NYPD was systematically ticketing legally parked cars to the tune of millions of dollars.

He was not intentionally trying to be antagonistic to NYC government or NYPD for that matter. He used Open Data and made a discovery he felt obligated to share. He had been trying to engage with NYPD for a while on this story, but with no response. That is why he reached out to me. He figured my team and I would be a useful conduit to city leadership on this matter.

The NYC Open Data team discussed this blog and ultimately decided that it was good for city leadership, New Yorkers and the mission of NYC Open Data and quite frankly, open data all over the world.

Ben was showing the value of it and demonstrating how open data was meant to be used; to give ordinary citizens insight into the operations of their government, making it transparent so they can learn about their city and engage with their government.

At this point I did what I could to engage with city hall and get them to talk with Ben about his blog before publishing. My team and I, and of course Ben, saw this as a good thing for New Yorkers. It was time for those of us (primarily me) who travel the world getting invited to conferences to extol the virtues and value of open data and how it is a powerful tool for transparency, to put up or shut up.

I won't bore you with any additional intricacies of this story except to share that soon after, Ben published his blog appropriately titled: *the NYPD was systematically ticketing legally parked cars for millions of dollars a year- open data just put an end to it.*

Open data sharing aids collaboration. During a data drill, Dr Mashariki identified 11 agencies who could integrate datasets on buildings in New York City – resulting in an intelligent platform for collaboration.

If you don't read the whole blog I want to share the end which was – and still is, in my opinion – as well a written piece about the value of open data that I have ever read.

NYPD released a statement, at the end specifying: "Thanks to this analysis and the availability of this open data, the department is also taking steps to digitally monitor these types of summonses to ensure that they are being issued correctly."

Ben wrote in response to NYPD's statement:

" This is what the future of government could look like one day. This is what open data is all about.

"This was coming from the NYPD, who is not generally celebrated for its transparency, and yet it's the most open and honest response I have received from any New York City agency to date. Imagine a city where all agencies embrace this sort of analysis instead of deflect and hide from it.

"Democracies provide pathways for government to learn from their citizens. Open data makes those pathways so much more powerful. In this case, the NYPD acknowledged the mistake, is retraining its officers and is putting in monitoring to limit this type of erroneous ticketing from happening in the future. In doing so, they have shown that they are ready and willing to work with the people of the city. And what better gift can we get from open data than that."

This is just one example of the power of an initiative that is still in its most nascent stage.

With respects to open data and building trust in government, I do think the best is yet to come.

Combat social problems with strategic, data-driven solutions

How location-based analytics made practical improvements to the lives of New York's citizens

Cities all around the world – from Sydney to Singapore, Kuala Lumpur to New York – are often faced with the same issues. How do you protect property renters from discrimination; engage parents in the education system; and prevent the large-scale spread of a deadly disease?

In my experience, the answer is usually the same: location-based data analytics.

Below are some practical examples of how this approach to analytics has enabled New York City to provide more equitable housing; prevent an epidemic; and accurately target communications to residents.

Preventing housing discrimination

New York's Human Rights Law prohibits landlords from discriminating against tenants. Yet income discrimination is one of the top housing-related complaints received by the city's Commission on Human Rights. It had "years upon years" of data from calls made by victims including the landlords' names and addresses.

In 2017, the city's analytics team worked with the commission to predictively identify where and which landlords were likely to discriminate. The team combined complaints data with crime, education, city planning and housing data, turning it into a map that geographically identified where these occurrences were likely to happen.

The commission then sent in actors and actresses posing as tenants to identify landlords who were discriminating. Some had housing vouchers and others didn't, to test how they were received by landlords and building management companies. In total, it conducted over 300 such tests.

Within two months, the commission charged five landlords and housing brokers – which together control about 20,000 units across the city – for repeatedly discriminating against tenants based on their income. The commission has also stepped up efforts to file investigations on behalf of the city, with the 120 cases filed in 2017, up from just 22 in 2014.

Preventing a legionnaires epidemic

In 2015, 128 people were infected and 12 people died as a deadly disease spread across the city. Legionnaires bacteria was being spread through untreated water in cooling towers on buildings. With over 1 million buildings in the city, tens and thousands of lives at risk, and limited resources for inspection, my team and I were asked to use data to identify locations with the biggest risks.

The city, however, had no existing list of all cooling tower locations. We worked round-the-clock for weeks, pulling in fragments of data from multiple agencies, to build a machine-learning algorithm to identify buildings likely to have contaminated cooling towers.

The team raised the hit rate for identifying cooling tower locations from 10% to 80%, with location-based data analytics.

Targeting resident communications

In 2018, New York Mayor Bill de Blasio announced a new scheme offering free pre-kindergarten for all three-year-olds in the city. To be a success, however, the scheme had to be inclusive of all people and have a high enrolment. The government launched a citywide campaign, sending officials door to door telling people about the new scheme.

But with close to 9 million people living in the city, officials had to be micro-targeted to ensure all eligible children had a chance to register. We needed to give every New Yorker the opportunity to know about and enrol in this program. We used data to identify parents who were likely to need and want free pre-kindergarten.
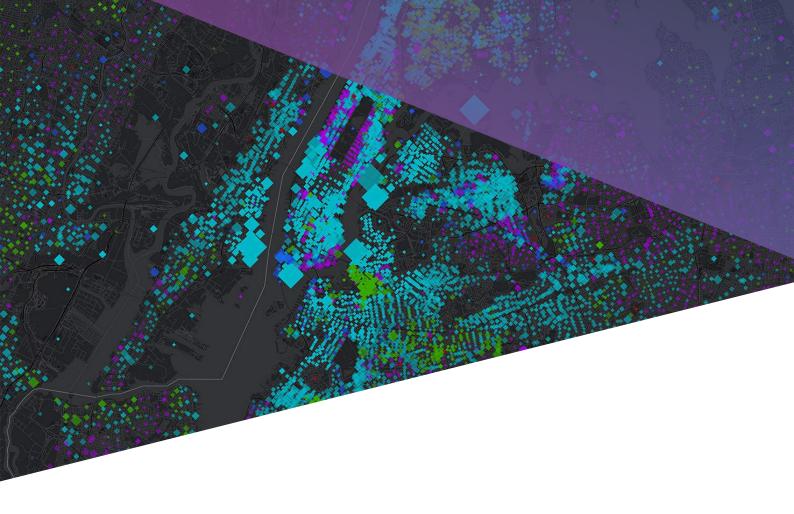
Using advanced location-based analytics to target our resident communications, we were able to increase enrolment by 117%.

With targeted and accurate information, frontline officials were able to better engage and educate parents about the program. This allowed people to pre-enrol before the start of the scheme.

The data accuracy also ensured residents had a strong customer experience with the government. People weren't being repeatedly bothered by different city officials and canvassers did not go knocking on the wrong doors. We changed how government was perceived by a lot of New Yorkers through that effort.

These case studies all show how vital analytics can be for a modern city. One of the things I've learnt is that everything in a city happens somewhere, so location is of utmost importance.

We've come a long way to make our cities smarter and more liveable, and we can still go further by using location-based analytics.

LESSON 4

Don't wait for existing systems to break down before implementing a modern city-wide data strategy

How to incorporate new data strategies into existing legacy frameworks

Some people believe you must first be able to manage data as an asset. Then, you can use it for operational purposes such as predictive analytics, operation business intelligence or strong city indicators and metrics.

Then there are some who are in the camp of doing analytics projects. They believe the more you do them, over time, the better you will get at managing the data. This is because you will be getting better at executing analytics work.

Instead, I believe that it makes more sense to do both in concert with each other.

The language of data analytics

We know it is easier to learn a new language when you are younger as opposed to when you are older and more set in your ways. The same thing can be said for adopting data as a citywide asset strategy.

It is exponentially easier to do this when you are in the "young" stage: buying your data management systems for a city from the ground up. Unfortunately, we don't have that luxury in many cities. In most, if not all our cities, we have legacy systems that do their job methodically and consistently, and may be difficult and costly to retire and replace.

In this instance, the challenge is how do we incorporate data as an asset strategy into the existing legacy city data framework? In many instances, these systems were not built and deployed with data sharing, or transparency, in mind.

Executing projects that have defined and discrete objectives requires sharing from multiple citywide and private-sector data sources. This process typically runs into obstacles, and these require both compromise and creative approaches to delivering solutions.

The challenges are akin to deploying a full immersion strategy for learning a new language. It is frustrating and seems impossible at first, but give it time and it becomes easier and easier to understand and then speak the language.

When it comes to learning a new language, it's also important to consider the learning tools, process and syntax. In the world of data analytics, this is directly analogous to understanding the quality of the technology we hope to deploy, as well as how we use this technology and when. The language syntax is similar to understanding and deploying data standards and protocols for security, privacy, storage and sharing.

Developing a proficiency in data

It takes time, meetings and countless conversations across the city to get consensus and buy-in for any data project. And not everyone is fluent in the language of data.

I would be remiss if I did not add that learning how to speak a new language differs based on the type of language. Understanding the origin of that language

is key. For instance, the etymology of the language may shift how you learn that language. I know that I will take on learning Chinese in a much different way than Spanish, based on understanding the culture and etymology of the language.

What I am getting at here, is that understanding the 'data etymology' – or base framework of the data a city uses to impact operational capabilities – requires an understanding that most of this data tends to be geospatial in nature.

Valuing city data as an asset

Building a capability to store this type of data as an asset requires the city to understand this location-based framework.

Cities must learn to use their data as an asset. This requires implementation and execution of operational analytics work. It also requires thoughtful data management best practices and tactics.

I don't have an opinion as to which should come first; managing the data or using it in a project.

I do, however, know that for a city to be prepared to use data when solving complex problems in a timely fashion with a high level of accuracy and precision, both have to be happening in parallel.

And location should be the etymology for this to happen.

Encyclopedias of organisation data are only as good as the people who analyse them

Why every smart city needs an analytics team: Pinpointing problems and targeted solutions

Cities are undergoing digital transformation with ever-increasing ways to capture data –from connected streetlights to CCTV to traffic flow and collision information transmitted from smart roads.

But many cities fall short of their digital potential because they focus on collecting and using that data in narrow, traditional ways.

A truly smart city goes beyond this, improving quality of life and the climate for business by analysing data in service of specific challenges and using it to make good decisions in real-time on a range of critical issues, from public safety to open data to economic investment.

Probing deeply into big data and pinpointing solutions to important problems demands a chief analytics officer (CAO) rather than a chief data officer (CDO).

Both are essential to a smart city or smart business, but they are quite different.

Why hire a CAO?

The company or city that understands the strategic capability that a CAO and their team can bring will gain an economic advantage through the ability to resolve pressing issues quickly.

By efficiently untangling problems, a company can become more profitable and a city safer and more liveable, prosperous and attractive to residents, tourists and new businesses.

It's not at all surprising that several recent articles indicate the number of CAOs in the private sector and the number of cities with analytics teams are starting to grow.

What capabilities should your analytics team possess?

There are no university degrees or majors dedicated to producing analytics officers. But we now know from experience what type of people you need to staff an analytics team – adept at going down unexplored paths in pursuit of answers.

In a crowded, complicated city like New York, where health, economic or social problems can quickly multiply, data analysis cannot become a years-long doctoral thesis. It often is a life-or-death matter that demands quick answers.

Getting things done in New York meant jumping in to attack problems that threatened residents' health, safety and quality of life, and analysing those issues on the run.

Members of a CAO team need to know how to talk with lawyers, police, firefighters, renters, landlords, doctors, nurses and city officials in all manner of departments and bureaus.

Those same people need to understand something of the ways that economics, social trends and politics can affect the problem, and be willing to explore new avenues of inquiry.

The people doing this work for cities or businesses – especially the leaders of analytics departments – must be curious, quick on their feet, able to speak the language of other experts, creative in their approaches, aggressive in their pursuit of the right data, and able to use location analysis to bolster their findings.

> When a smart city assembles such a team of problem solvers, the result is improved efficiency and money saved for taxpayers.

In the case of the NYC firefighters and the illegally renovated buildings, these analytical efforts helped save lives and prevent injuries. I'm proud of that work. I know that many other cities can create similarly skilled teams and generate similar results.

How can you build your analytics team?

In addition to the curious, aggressive, flexible people I talked about above, smart cities really do need a strong data team, which begins with a good chief data officer (CDO). A CDO doesn't compete with a chief analytics officer. They complement each other.

You can't pull great answers out of poorly organised data.

Your CDO, whether in the public or private sector, needs to make sure that the categories of data fit the organisation and are readily accessible to those searching for answers to urgent problems.

Citywide problems will cut across departmental lines, so the easier the access and the wider the data sharing, the better the outcomes.

Most companies and cities have huge stores of data, but that information may simply sit there, like volumes of neatly stacked encyclopedias.

A skilled analytics team, led by a chief analytics officer, can draw on that data to improve efficiency in solving problems and strengthen decision-making.

To find out more, visit esriaustralia.com.au/smartcities

esri Australia

THE
SCIENCE
OF
WHERE™

Contact Esri Australia

For information on how to apply GIS solutions
in your business, call us on 1300 635 196 or
visit **esriaustralia.com.au**

 twitter.com/esriaustralia   f facebook.com/esriaustralia
 connect@esriaustralia.com.au   ⊕ **esriaustralia.com.au/blog**

THE
SCIENCE
OF
WHERE™