

The Analytics Playbook for Cities

A Navigational Tool for Understanding Data
Analytics in Local Government, Confronting
Trade-Offs, and Implementing Effectively

Amen Ra Mashariki and Nicolas Diaz
David Eaves, *Series Editor*

August 2020



The Analytics Playbook for Cities

A Navigational Tool for Understanding Data
Analytics in Local Government, Confronting
Trade-Offs, and Implementing Effectively

Amen Ra Mashariki and Nicolas Diaz
David Eaves, *Series Editor*

August 2020



ABOUT THE AUTHORS

Amen Ra Mashariki is the Global Director of the Data Lab at WRI. There he works across programs, international offices and centers to identify data solutions to turn big ideas into action to sustain our natural resources—the foundation of economic opportunity and human well-being. Amen Ra is also a Data + Digital fellow at the Beeck Center for Social Impact and Innovation where he partners with academic, public and private sector thought leaders to shape best practices and strategies on how to use data to maximize impact in our communities.

Prior to this, Dr. Mashariki has served as Head of Machine Learning at Urbint, adjunct faculty at NYU’s Center for Urban Science and Progress (CUSP), Fellow at the Harvard Ash Center for Democratic Governance and Innovation, Head of Urban Analytics at Esri, Chief Analytics Officer for the City of New York, Director of the Mayor’s Office of Data Analytics at the City of New York, White House Fellow, and Chief Technology Officer for the Office of Personnel Management of the United States.

Amen earned a Doctorate in Engineering from Morgan State University, as well as a Master and Bachelor of Science degree in Computer Science from Howard and Lincoln University respectively.

Nicolas Diaz Amigo is a graduate from the Master in Public Policy program at Harvard Kennedy School. He has had research and fellow roles at the Bloomberg Harvard City Leadership Initiative and at digitalHKS. Professionally, he served as the first Coordinator of Public Innovation at the Mayor’s Office in Santiago, Chile, City Hall and has consulted for cities and local government organizations across the world on the topics of public sector innovation, data analytics, digital transformation, performance management, budgeting, and process improvement.

ABOUT THE SERIES EDITOR

David Eaves is a lecturer of Public Policy at the Harvard Kennedy School where he teaches on digital transformation in Government. In 2009, as an adviser to the Office of the Mayor of Vancouver, David proposed and helped draft the Open Motion which created one of the first open data portals in Canada and the world. He subsequently advised the Canadian government on its open data strategy where his parliamentary committee testimony laid out the core policy structure that has guided multiple governments approach to the issue. He has gone on to work with numerous local, state, and national governments advising on technology and policy issues, including sitting on Ontario’s Open Government Engagement Team in 2014–2015.

In addition to working with government officials, David served as the first Director of Education for Code for America—training each cohort of fellows for their work with cities. David has also worked with 18F and the Presidential Innovation Fellows at the White House providing training and support.

ACKNOWLEDGMENTS

This report has been possible with the sponsorship of digitalHKS, a project at Harvard Kennedy School that studies the intersection of digital technologies and governance. The authors are grateful for the generosity of the public servants who provided case studies. Additionally, many individuals provided invaluable feedback to earlier version of this report, including Kadijatou Diallo, Tommaso Cariati, Westerley Gorayeb, Natasha Japanwala, Lauren Lombardo, Michael McKenna Miller, Nagela Nukuna, Emily Rapp, Naeha Rashid, Imara Salas, and Blanka Soulava. This report is an independent work product and views expressed are those of the authors.

THE ANALYTICS PLAYBOOK FOR CITIES

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

ABOUT THE ASH CENTER

The Roy and Lila Ash Center for Democratic Governance and Innovation advances excellence and innovation in governance and public policy through research, education, and public discussion. By training the very best leaders, developing powerful new ideas, and disseminating innovative solutions and institutional reforms, the Center’s goal is to meet the profound challenges facing the world’s citizens. The Ford Foundation is a founding donor of the Center. Additional information about the Ash Center is available at ash.harvard.edu.

This research paper is one in a series published by the Ash Center for Democratic Governance and Innovation at Harvard University’s John F. Kennedy School of Government. The views expressed in the Ash Center Policy Briefs Series are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. The papers in this series are intended to elicit feedback and to encourage debate on important public policy challenges.

This paper is copyrighted by the author(s). It cannot be reproduced or reused without permission. Pursuant to the Ash Center’s Open Access Policy, this paper is available to the public at ash.harvard.edu free of charge.

CONTENTS

Executive Summary	1
1. Introduction: The Potential Use and Misuse of Data Analytics in City Government	2
2. Definitions	5
2.1. What an Analytics Team Is	5
2.2. What an Analytics Team Is Not	5
3. Principles	7
4. The Plays	9
4.1. Setting Up the Right Team for Success	9
4.1.1. Step Zero: Making the Case for Data Analytics	9
4.1.2. Defining Structure: Decentralized or Centralized?	10
4.1.3. Who to Hire	12
4.1.4. How to Hire	13
4.2. Starting from Scratch	15
4.2.1. How to Find Your First Project	15
4.2.2. Making the Case for More Funding	16
4.3. The Analytics Team Day-to-Day	17
4.3.1. Understanding the Multiple Uses of Analytics: Repertoire of Actions	17
4.3.2. Picking the Right Projects	19
4.3.3. Working with Other Departments	20
4.3.4. Data Stewardship	21
4.3.5. Pushing Projects to Production	21
4.3.6. Branding and Communications	22
4.4. Important Considerations	23
4.4.1. Open Data	23
4.4.2. The Pitfalls of Analytics: Privacy, Security, and Algorithmic Bias	24
5. Summary: Data Analyst, Go Out There and Listen	27
6. Annex	28
6.1. Case Study: Chicago's CDO Has a Centralized Mandate	28
6.2. Case Study: Piloting London's Office of Data Analytics	29
6.3. Case Study: Boston's Implementation of a Centralized Data Warehouse	31
6.4. Case Study: Transport for London and Leveraging Data Products	32

EXECUTIVE SUMMARY

Properly used data can help city government improve the efficiency of its operations, save money, and provide better services. Used haphazardly, however, the use of analytics in cities may increase risks to citizens' privacy, heighten cybersecurity threats, and even perpetuate inequities.

Given these complexities and potentials, many cities have begun to install analytics and data units, often head by a chief data officer, a new title for data-driven leaders in government. This report is aimed at practitioners who are thinking about making the choice to name their first CDO, start their first analytics team, or empower an existing group of individuals.

In the following pages, we provide city leaders with:

- **Definitions.** To understand what an analytic team in city government looks like, as well as presenting alternative team structures in the realm of data, digital government, and innovation.
- **Principles.** The attitudes and mindsets that are useful when trying to enact ambitious change in local government, such as *ensuring executive buy-in* and *seeking allies outside of government*.
- **Plays.** The organizational, strategic, and tactical considerations that a city must think about for starting and supporting an analytics team. These are presented in a sequential order: how to set up a team for success, building up from scratch, managing the day-to-day, as well as important considerations (like privacy and ethics) that should be present throughout.

1. INTRODUCTION: THE POTENTIAL USE AND MISUSE OF DATA ANALYTICS IN CITY GOVERNMENT

Imagine a world where our cities not only deliver the services that we expect from them, from weekly trash collection to running our schools to supporting local businesses, but also tailor those services to each of its citizens. Like Netflix recommending the next show you should watch based on your previous views or Amazon steering you to a product you didn't even know you needed, local governments could connect citizens who request a service with other social programs they qualify for.

Imagine a city hall not only reacting to problems—a student dropping out of school or a building catching on fire—when they occur, but using data to identify risks and take preventive measures before they happen.

These are not pipe dreams. Municipal departments already possess an abundance of data, from utility bills to tax records to school enrollment histories, though the value of that data is often locked away in silos. **City analytics, a discipline that combines operations research and data science, has the potential to unlock the big insights behind the big data** and move local governments toward higher efficiency and effectiveness in the provision of services and the prevention of social ills.

When it comes to the use of advanced analytics, the public sector needs to catch up to the progress seen in the private sector over the past few decades. But strictly copying private sector practices is not the right approach. There are considerations for dealing with data collected by a public institution. With all the potential analytics hold for improving the delivery of local services, its risks must also be carefully examined by analysts, the senior leaders who oversee their work, and civil society at large. **A misuse of analytics may pose cybersecurity threats, threaten privacy, or reinforce existing inequities.**

Lessons from NYC: Breaking silos increased the need for cybersecurity

To exemplify the risks of misusing analytics, let's take the example of cybersecurity. A few years ago, New York City was considering improving its monitoring system for underground assets such as sewage pipes that were being tracked by several agencies, each with its own software system and database. Bringing all of those together into a unique system would have allowed for more efficient management. But while a breach of any particular databases would not be particularly catastrophic, having the city's entire underground mapped by bad actors would be a major threat. The standards for how safely data is stored and shared internally had to be reexamined.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Doing analytics in government also comes with unique barriers, such as the difficulty of hiring individuals with the right technological background or moving an organizational culture that does not always welcome innovation.

But that doesn't mean there is no path forward. In fact, an increasing number of cities are beginning to systematically incorporate analytics, often in the form of new teams headed by leaders with titles like chief data officer, chief analytics officer, or chief performance officer. This playbook examines successes and failures, and shows what hard choices have so far been made along the way. **An analytics team must face difficult questions that do not have clear answers: Who do we hire first? How do we prioritize our projects? How much time should we spend communicating our successes?** By illuminating some of the trade-offs cities have faced, we hope this document will help city officials find a strategy that works for their own context, limitations, and objectives.

In sports, it is important for a coach to understand the wide variety of plays that may come in handy in different circumstances. A playbook provides a sense of orientation in a context of high complexity. What should happen at the start of the game, what do we do when we are in overtime, etc. But just like in sports, understanding the generic plays is not enough to win the game. It's all in the execution and flexibility in adapting to whatever resources are available.

The lessons in this playbook rely heavily upon Amen Ra Mashariki's experiences as head of the Mayor's Office of Data Analytics (MODA) in New York City, where he led a team of nine analysts and policy specialists. To illustrate some of the lessons throughout this document, we will call out these insights in boxes like the one on the previous page. But we have also put these ideas to the test by consulting some of the leading professionals in the field, who have generously provided case studies that illustrate the potential and barriers of urban analytics.

One thing readers should be wary of is the temptation to think that every problem can be solved with data and technology. Using analytics heedlessly may even make some problems worse. The intention of this report, therefore, is to allow for a critical application of sophisticated tools. We will address several dimensions of those tools including **learning how to identify which problems are suited for an analytics approach, understanding potential issues, setting up the right team to tackle those issues, and making sure the team gets the right resources for success.**

This document is structured as follows. We start by defining what an analytics team is and what it is not. Then we highlight principles that guide successful teams.

THE ANALYTICS PLAYBOOK FOR CITIES

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Finally we will highlight specific plays: the key trade-offs, questions and frameworks that an analytics team in local government must consider in order to effectively pursue its mission.

This playbook won't give you all the answers for starting an analytics team. Instead, it is are meant to orient you toward questions that allow you to connect the work you want to do with whatever is important and feasible right now. The skill sets of your employees, the legal framework you are operating under, and even the political cycle will mold your journey. We have tried to offer some thoughts and ideas, but ultimately you will have to decide which plays to focus on at each time. To get better at the game, you have to start somewhere.

2. DEFINITIONS

In their desire to move toward data-driven organizations, well-intentioned public managers may fail to fully grasp what they are getting themselves into. **Not all organizations need an analytics team, but by** providing a clear sense of what analytics teams are and what they are trying to accomplish, we hope to make it easier for senior officials to choose whether to pursue this path.

2.1. What an Analytics Team Is

Analytics teams may take different names and vary in size, scope, and mission across different jurisdictions.¹ Our basic definition of an analytics team in local government is a specialized group that applies data analysis to uncover insights with the hope of improving the operations of city departments.

In practice, what we most often find is a team between one and 20 employees that has been formally integrated in the bureaucracy. This team will leverage data-intensive tools such as machine learning or mapping to define, evaluate, and propose solutions to public problems. Its mission statement is usually some variation of “improving the outcome of what the city does to create more public value for constituents at lower cost”—though the most pressing areas to tackle to deliver public value will change depending on where you are and whom you ask.

2.2. What an Analytics Team Is Not

Other data-driven management and digital-government adjacent groups can take these forms:

- A traditional information technology (IT) team that maintains technology assets and keeps systems running
- A performance management team that works across departments to define the relevant metrics for success
- A digital service team that ensures interoperability of the data, such as allowing members of one department to easily access data from other departments
- An open data team that maintains an online portal aimed at sharing city data sets with the general public
- A data governance team that sets citywide standards and convenes analysts across the city to build a citywide data culture

.....
1 Jane Wiseman. (2017). Lessons from Leading CDOs: A Framework for Better Civic Analytics.

.....

THE ANALYTICS PLAYBOOK FOR CITIES

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Note that these functions can be complementary to the work of an analytics team. In some places there may be overlap (for example, a chief data officer may set open data policies), and cities may choose to prioritize some of these over others.

3. PRINCIPLES

While no single blueprint can work for every city, you will find throughout this report some common principles that underlie any effort to bring more analytical thinking to local governments. Teams should:

- **Commit to better service delivery.** An analytics team must be driven by a desire to make government work better for its citizens in a transparent and accountable way.
- **Be both the disrupter and the listener.** Given the nascent state of analytics in government, analytics teams may have to act like internal entrepreneurs who question the way things are done. But this desire for disruption must be paired with a profound curiosity to understand the nuances and challenges that led to the status quo.
- **Start with the problem, not the solution.** An analytics team should avoid the temptation to run around with fancy tools in search of places to apply them. Instead, its strategic outlook and tactical moves should be centered around pressing challenges faced by the city or its individual departments. And when a need is found, a sense of practicality and viability should supercede the drive for sophistication. Simple is better than complex.
- **Ensure executive buy-in.** As talented and persuasive as data-driven public servants may be, support from key stakeholders at the top is essential. Whatever your specific institutional arrangement, buy-in from your mayor, her chief of staff, the city manager, or someone sufficiently high up will allow you to overcome the resistance to change you'll encounter along the way.
- **Work in the open.** The work should not be top secret. The projects undertaken should be considered part of a conversation about public service delivery that is appropriately communicated within city hall and with the public. Blog posts can help explain the aim of each project and code can be posted online for scrutiny. A commitment to transparency should ensure that those seeking information about a process or its results can easily access it.
- **Find allies.** You will require allies both inside and outside the organization. In this playbook we will discuss in detail ways you can engage with department heads to create quick wins, when to extend ties to academia, and how the press can come in handy.

THE ANALYTICS PLAYBOOK FOR CITIES

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

- **Quickly deliver value, but think long term.** The team will need to rapidly demonstrate what it is bringing to the table and deliver quick wins. However, team members must also be aware that to create even more value they need to think about long-term investments in infrastructure and changes to the work culture. An effective manager must carefully consider both time frames simultaneously.

4. THE PLAYS

4.1. Setting Up the Right Team for Success

This section begins by looking at the most relevant part of your analytics effort: the people behind it and how they should be organized.

4.1.1. Step Zero: Making the Case for Data Analytics

How do you argue the need for an analytics team?

Before beginning to establish an analytics team, stakeholders may need to be persuaded. Whether it is a city administrator, a councilwoman, or the general public, a convincing case must be made that a new team will help the city reach its goals.

Jane Wiseman, CEO of the **Boston-based management consulting firm** Institute for Excellence in Government, in studying cities investing in analytic efforts,² points toward multiple types of value that can be attributed to data teams:

- Fraud detection cost savings
- Efficiency improvements that reduce costs
- Accuracy improvements that reduce costs
- Increased revenue capture
- Efficiency improvements that improve outcomes
- Operational changes that increase safety
- Increased faith in government due to more transparency

Naturally, not all stakeholders will care equally for every one of these. A city administrator may be more concerned with cutting costs than with transparency, for example. It is important to **tailor the message in a way that considers particular stakeholder interests**.

On the issue of improving city finances, a McKinsey study³ on the use of data analytics in government to find fraud found that the return of investment could be as high as 15 times the cost. Wiseman also points toward two municipal data teams that self-reported their return on investment (ROI): The Louisville Metro Government calculated a five-to-one return for analytics efforts and the Cincinnati Office of Performance and Data Analytics calculated a \$6.1 million added value to the city in two years of operation, a nine-to-one return.

.....
² <https://datasmart.ash.harvard.edu/news/article/case-government-investment-analytics>.

³ Susan Cunningham, Mark McMillan, Sara O'Rourke, and Eric Schweikert, "Cracking down on government fraud with data analytics," October 2018, McKinsey & Company.

4.1.2. Defining Structure: Decentralized or Centralized?

Should the analytics team be in one department or agency or spread out among several?

An analytics team may be a single central office that takes on projects across the city (centralized model) or a collection of employees who are spread among various departments (decentralized model).

To decide between the two models, a useful rule of thumb is to **think about what types of projects would generate the most immediate value for senior sponsors.**

If you want to use data to tackle an issue that is handled by a single department (e.g., crime prevention, which is the responsibility of the police department), then **a decentralized model, with analysts that sit permanently in that department, will allow for more sophisticated and in-depth research.**

On the other hand, if your aim is to address a challenge where different agencies share responsibility (e.g., you want to know which buildings to target for inspection in a city that splits code enforcement tasks among the fire department, housing, etc.), then you should establish a centralized, dedicated unit. **Having analysts who work with multiple departments will avoid task duplication and also reap the benefits of breaking down data silos** instead of having each department create from scratch its own models with partial data.

In “Data-Driven Government: The Role of Chief Data Officers,”⁴ published by the IBM Center for the Business of Government, Wiseman offers a systematic comparison of the centralized and decentralized models for government agencies, as seen in the following table.

.....
4 http://www.businessofgovernment.org/sites/default/files/Data-Driven%20Government_o.pdf.

.....

THE ANALYTICS PLAYBOOK FOR CITIES

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

	Centralized	Decentralized
Chief data officer (CDO) focus	<ul style="list-style-type: none"> Analytics resources in a single team under the CDO's leadership CDO team provides data and analytics services to key executives and managers, functioning as an internal consultant Team works in partnership with executives and managers to define scope and project needs Some bureaus or agencies—typically the better resourced or statistics-driven ones—may have their own analytics resources, but the majority rely on the centralized CDO team 	<ul style="list-style-type: none"> CDO team creates distributed capacity across government by embedding talent in bureaus through training and coaching CDO team creates tools, platforms, and data standards that speed adoption of data skills in bureaus Each bureau/agency/office is responsible for developing its own analytics capability, which can range from budget and policy analysts who can complete basic descriptive statistics and dashboards to analysts with the skill to perform data science tasks such as predictive analytics
Ideal for	<ul style="list-style-type: none"> Specialized skills or subject matter expertise needed for highly technical work Analytics projects that require a high degree of confidentiality, such as investigations 	<ul style="list-style-type: none"> Large-scale enterprises with similar or low-complexity operations spread across many bureaus, divisions or geographies Agencies with a high level of existing data skill or broad adoption of data literacy, such as scientific or statistical agencies
Benefits	<ul style="list-style-type: none"> Centralized pool of analytics talent allows sharing of specialized skills across the enterprise from a common hub, which saves money since highly trained analytics staff can be expensive for government Efficiencies are gained via peer support and collaboration among team members Centralized team is better able to standardize tools and processes across government, which can save time and money and help develop deeper expertise in the chosen tools and methods Team can facilitate cross-organizational data initiatives due to its enterprise-wide view of data assets and needs 	<ul style="list-style-type: none"> Leaders and managers in bureaus have more control over their analytics resources, may get more timely responses to their requests, and may also more immediately deploy analytics insights Analysts embedded in bureaus develop subject matter expertise that makes them valuable to their leadership and speeds time to results Embedded analysts can foster greater adoption of data culture across enterprise, which can lead to faster organizational culture change Skills gained in self-service analytics are transferable across government, spreading benefit
Limitations	<ul style="list-style-type: none"> Slow growth of sustainable analytics talent in the bureaus Can be challenging to achieve scale with a small centralized team, as surge capacity may need to be deployed for a high-priority task 	<ul style="list-style-type: none"> Putting decision-making and control of analytics in the hands of bureau heads leads to uneven attention and results, with some investing heavily and others giving it low priority or not appointing an analytics officer unless compelled to do so Decentralized model limits peer cohorts for data-focused employees and may result in a more limited career ladder

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Case Study: Chicago's CDO has a centralized mandate⁵

The first U.S. city to have a chief data officer (CDO) was Chicago. After briefly being part of the mayor's office, the position was then moved to the Department of Innovation and Technology.

Tom Schenk Jr. served as the city's second CDO and as deputy director of IT, which gave him both a centralized mandate to drive data analytics and the operational responsibility for maintaining and upgrading the city's databases and digital platforms.

This centralization of responsibilities allowed Schenk and his team not only to push for concrete data analytics projects, but also to take control of data governance. They began building a data inventory of the entire city and optimizing the data infrastructure for easier analytics in the future.

Being able to inventorize the data was essential for setting up future successes. However, Schenk warns that quick wins are an absolute necessity. "Stuff will get hard and you will need to ask a lot of people [for help]," he says. "An inventory of data will take a lot of time and effort, and both residents and internal stakeholders will not perceive its value until you show some progress and real, tangible results."

Schenk and his team also had to consider how to choose which projects to spend time on. To decide, the team worked closely with the University of Chicago, creating a framework to screen projects and assess the data maturity required for meaningful analysis in conjunction with the master of science program in Computational Analysis & Public Policy.⁶

A longer write-up of this case can be found in the annex.

4.1.3. Who to Hire

Who is the first person you hire for your analytics team? What about the second, third, and fourth?

There is no predetermined profile or career progression for the team members in an analytics office. One reason for that is the need for teams that cover a wide variety of tasks, from project management and performing complex analyses to knowing how to build effective cooperation with the domain experts and more.

An article published in the [Harvard Business Review](https://hbr.org/2019/03/why-data-science-teams-need-generalists-not-specialists) in 2019 noted that data science teams in the private sector often attempt to discover "profound new business capabilities," and because of this such **teams must be made up of generalists who are focused on learning and iteration rather than specialists who do only one thing efficiently.**⁷ This argument can be made even more strongly for local governments because of the immense quantity and complexity of the services they provide.

.....

5 Interview with Tom Schenk (October 2019).

6 http://www.datasciencepublicpolicy.org/wp-content/uploads/2018/05/Data_Maturity_Framework_4.28.16.pdf.

7 <https://hbr.org/2019/03/why-data-science-teams-need-generalists-not-specialists>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

For your first team member, consider looking for someone within the organization who has some local experience, rather than hiring out of a prestigious economic-analysis consultancy or data science firm. Even if this is a job that will require a good understanding of analytical methods, the first employee's main challenge at first will be effectively navigating the tacit idiosyncrasies of a political and organizational environment. An urban planner at the transportation department, for example, may have better insight into both the current challenges the city is facing and whom to partner with to get stuff done. Keep your eyes open.

When considering further hires, around analytics offices across the world, it is common to find professions as diverse as:

- Data scientists
- Urban planners
- GIS managers and analysts
- Political scientists
- Social science researchers/analysts
- Mathematicians and statisticians

Some skill sets that will be valuable regardless of previous occupation will be empathy (to understand the challenges of others), communication (to work effectively in collaborative teams), and curiosity (to always be looking out for status quo thinking that needs to be challenged).

Keep in mind that cultural diversity adds to the strength of the team (especially if it mirrors the diversity in your city) and will help avoid analytic projects with blindspots in issues such as race, differences among neighborhoods, etc. For more on this, see the passage on algorithmic bias in section 4.4.2, “The pitfalls of analytics.”

4.1.4. How to Hire

How do you overcome some of the challenges in attracting talented and motivated individuals to local government?

Once you start looking outside the organization, it will not be easy to hire who you want. Young, ambitious professionals well-versed in analytical skills may feel more at home in a small start-up than in a big, clunky bureaucracy. A McKinsey article about trends in the workforce⁸ highlighted the importance millennials put on flexibility, the availability of mentor relationships, and the autonomy to tackle their own projects—qualities not usually associated with the public sector.

.....
⁸ <https://www.mckinsey.com/business-functions/organization/our-insights/millennials-burden-blessing-or-both>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Moreover, there is the issue of money; usually the public sector has strict rules about how much to pay people according to their place in the organization chart, and the average starting salaries for data scientists is often higher than what most agency heads earn—so highly skilled professionals will be offered significantly less than what they could earn elsewhere.

One advantage that you do have is the ability to offer impact and purpose. You can structure your portfolio of projects in a way that gives each analyst the opportunity to work on complex urban challenges that may affect the lives of thousands of citizens, while bringing forth new methodologies and ideas. Team members may also have access to high-level individuals in the city. But this also requires you to create a workflow with the proper autonomy for everyone on your team.

Local universities and colleges are an important resource. Whether you are located in a small city or a large metropolis, there should be a college nearby with a data science program. A few universities even offer specialized programs, such as the University of Chicago’s Master of Science in Computational Analysis and Public Policy⁹ and New York University’s Urban Analytics track for its Master of Urban Planning.¹⁰ At the very least, your local educational institution should have a statistics program. If searching locally bears no fruit, you might look into specialized networks for connecting the public sector with academia such as the MetroLab Network¹¹ and the Data Science for Social Good Fellowship at Carnegie Mellon University.¹²

Lessons from NYC: Accepting a faster changing team

At times, the traditional model of structuring a team in local government may need to be reconsidered. In New York City, Amen knew that if he hired young people he would not be able to retain them for long. So he kept a team with high turn-over and a frantic schedule of projects, offering his team the ability to work with autonomy on high-stakes issues.

Effectively engaging educational institutions means giving them access to interesting data sets and projects. For professors and students, getting their hands on real data for analysis is incredibly enticing. Graduate students must often do capstone projects with real-world clients in order to graduate. Finally, you may use this as an opportunity to create a pipeline for talent, offering student internships that may turn into full-time jobs by graduation. Regardless of the specific approach you take, **hiring**

.....
9 <https://capp.uchicago.edu/>.

10 <https://wagner.nyu.edu/education/degrees/master-urban-planning/urban-analytics>.

11 <https://metrolabnetwork.org/>.

12 <http://www.dssgfellowship.org/>.

should be a continuous task. You may have to be constantly looking at resumes, nurturing your relationships with academia, going to classes to meet with students, looking for interns, etc.

4.2. Starting from Scratch

Once in place, your analytics team will need to get started quickly. This will mean sorting out difficulties and finding a way to demonstrate value quickly. This section goes in depth into those crucial first steps.

4.2.1. How to Find Your First Project

How do you spot a “minimum viable product”?

Your first analytics project is important. It will provide an opportunity to prove the value of the team to leadership and to senior officials throughout the city, so you should start not with a solution in mind but with a real problem that the city is facing—ideally, a challenge that has been clearly identified as a priority by leaders.

To further build the case for the analytics team, it should be a problem that can be tackled by integrating data sets from across multiple agencies. This will show the value of breaking data silos and will probably uncover problems of poor data collection. Furthermore, problems where the data is scattered are often complex policy issues that require deeper thinking—something that may have been ignored so far.

Finally, this cannot be an intellectual exercise only. Although there is plenty of value in visualizing a problem that may have been hidden from the view of stakeholders, to truly cement your first project as a success you should work with your partners toward a clear strategy for delivery. Do not assume that this will happen on its own. Talk with partnering departments beforehand to clarify how analytical results can help with the implementation or piloting of operational innovations.

From the very beginning, identify partners across the entire city government who see the worth of your work. They may be senior officials or new employees. Regardless, make them your data champions. Empower them. But do not forget to communicate to the top; the senior advisor or chief of staff to the mayor may be very busy in her day-to-day, but she should still be frequently updated on what you are doing.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Case Study: Piloting London's Office of Data Analytics¹³

In 2016 a pilot began for the London Office of Data Analytics (LODA). It was orchestrated by the Greater London Authority; 12 of the 36 London boroughs; Nesta, an innovation NGO; and ASI, a data science firm.

LODA found its first project by asking for suggestions from the participating boroughs. Then, a workshop session was conducted to collaboratively assess the projects based on:

- Money saving potential
- Availability of data
- The ability to produce insights and delivering results within two months
- The ability to solve the problem without personal data sets

After the issue selection process, the pilot focused on one use case: leveraging predictive analytics to identify multiple occupations in houses that did not have the appropriate licenses. By combining data sets from multiple sources, the team sought to point inspectors toward infractors.

Ultimately the project was unable to produce meaningful insights, but it did help inform a protocol for data-sharing among the boroughs.

A longer write-up of this case can be found in the annex.

4.2.2. Making the Case for More Funding

How do you argue for more money?

Every budget appropriation process is different, as the financial, political, and technical idiosyncrasies play out. However, the more clarity there is in the analytics team's mission, the easier it will be to iterate that mission and construct a compelling narrative. It helps if the mission is closely aligned with the intentions of senior leaders. Often, analytics teams get stuck trying to do a little bit of everything, which muddles the argument for recurring funding.

You should be aware of the downside to following performance indicators, though, since important outcomes are often not measured. Performance indicators will usually point you toward things that are already being done by each department, potentially blinding you to exploring solutions that may not fit neatly within the existing bureaucratic structure. On the other hand, focusing on improving upon existing solutions may make the potential implementation more straightforward. This is a tension that you will have to navigate constantly.

.....

¹³ Sources: <https://londondatastore-upload.s3.amazonaws.com/LODA%20pilot%20report.pdf>, <https://data.london.gov.uk/loda/>, <https://data.london.gov.uk/city-data-analytics-programme/>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Once you have finished your first project you may choose to scale your work in either vertically, by growing the complexity of the analysis you provide to your partners (i.e., moving from describing trends in data sets to predictive work to prioritize resources), or horizontally, by growing the number of agencies with which you are collaborating.

Lessons from NYC: Look at performance indicators

A logical alternative would be to make sure the projects that the team implements are tied to performance indicators. For example, when Amen was leading the New York City team, they only worked on projects that were part of the yearly quantitative goals that had to be reported to the mayor's office and to the city at large. If a department had a proposition, the analysis had to have a logical and explicit way to move the numbers toward a goal. This discipline made it easier for the analytics team to track its impact in terms of city priorities met and dollars saved, thus making the case for growth.

4.3. The Analytics Team Day-to-Day

As the analytics team begins to find a comfortable place in the organization, it should consider how to transition into a sustainable workflow that delivers the most value. This section helps to illustrate what you can expect the team's day-to-day to look like. What sort of projects can the team offer? How does it prioritize projects? How should other departments be brought into the fold? And, how do you shift from merely reacting to thinking about the long term?

4.3.1. Understanding the Multiple Uses of Analytics: Repertoire of Actions

What can you use analytics for?

Once an analytics team is in place and has one or two projects under its belt, it should begin to look for other ways it could add value to ongoing projects, and how to let other departments understand what services the team can offer to help them do their work.

The table below is adapted from work performed by New York City's Mayor's Office of Data Analytics, and includes the categories of data analysis projects and real-use cases for a variety of city needs.

.....

THE ANALYTICS PLAYBOOK FOR CITIES

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

	Why would we want to do it?	Example application
Prioritizing Where to go first?	Ranking a list according to certain criteria can enable more efficient use of resources. Useful when [getting to the worst things]? earlier can mitigate potential negative effects.	To assist the Department of Education's work to make all schools ADA-compliant, MODA used DOE data to prioritize which schools to renovate in order to reduce the number of students with disabilities who needed to use buses.
Scenario Analysis What if?	Considering alternative events and their possible outcomes can help policymakers find the best course of action and plan for a greater range of possible policies.	As part of the Mayor's Office of Long Term Planning and Sustainability's research for a new commercial composting policy, MODA predicted how much waste local businesses would generate under various regulatory thresholds.
Anomaly Detection What is out of the ordinary?	Some processes can be improved by identifying and investigating outliers. Useful when looking for the exception is more feasible than examining every case.	Registration records of all kinds may have a number of files that display unusual characteristics. Flagging and examining those records may reveal procedural oversight or fraudulent transactions.
Matching What goes with what?	Matching can optimally pair two groups against a certain set of constraints. Useful for equitably distributing limited resources.	When the appointment scheduler for IDNYC was backlogged with duplicate requests, MODA helped match applicants to times and locations based on indicated preferences.
Estimating How much will a project cost?	Projects can be planned more effectively when time, materials, and costs are estimated in advance. Useful for quantifying the costs and benefits of new programs.	MODA worked with the Department of Housing Preservation and Development to estimate the resource requirements and program outcomes for a new set of Enhanced Contractor Review procedures.
Targeting Where to look?	Targeting can narrow an operational domain to enable better resource allocation. Useful for identifying a subset for a specific intervention.	MODA created a model to help identify buildings that have displayed a pattern of unsafe living conditions. This enabled the Tenant Harassment Prevention Task Force to follow up with inspections and enforcement actions when necessary.

4.3.2. Picking the Right Projects

How do you choose what to work on?

At some point, you will have to choose between multiple projects that are asking for the analytics team's resources and attention. Look for projects with the right:

- **Partners:** Do you have buy-in from stakeholders to try new solutions and put insights into practice?
- **Data:** Is there meaningful data that could lead to insights?
- **Impact:** Will the analysis help illuminate a solution for a problem that is relevant?

Chicago's approach was codified in a form given to technical departments that requested help from the analytics team,¹⁴ allowing users to quickly assess a number of alternatives. Not every city necessarily needs to develop its own questionnaire, as the day-to-day of project intake may be more messy, but there should be some sense of which projects advance the analytics team's mission and the overall goals of the administration.

		Center for Data Science & Public Policy THE UNIVERSITY OF CHICAGO			
Data Maturity Framework Questionnaire					
Scorecard Category	Question				
Problem Definition					
Problem Definition	What is the problem you are trying to solve? What does success look like/how much does the needle need to move?				
Interventions	What interventions do you have available to solve the problem?				
Impact	If this is successful, what impact will this project have? Will it encourage future projects/goodwill?				
Available Data	What data sets do you have access to relevant to the problem?				
Data Fields	What fields are in each of the data sources? (See Data Sources Worksheet)				
Size	How many people/addresses/facilities/entities does the data contain?				
Target Population	For this problem, what % of entities are at risk or have resources to be intervened?				
Data Governance					
Ownership	For the data sets that you have access to - do you own the data? Do you have permission to use the data? If you do not own the data, do you have the relationships with the data owner?				
Physical Accessibility	Is the data accessible outside the department/agency? Is there a VPN?				
Security Policy	What security policies and considerations need to be in place for each of the data sources? (HIPPA, FERPA)				
Implementation and Maintenance					

A portion of Chicago's data maturity questionnaire, developed alongside the University of Chicago, that provides a standardized approach.

¹⁴ http://www.datasciencepublicpolicy.org/wp-content/uploads/2018/05/Data_Maturity_Framework_4.28.16.pdf.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Look for performance metrics. If, prior to the collaboration, there is some measurement of whatever problem is being solved or service is being improved, that will help make the case for the value of the team in the form of dollars saved, extra customers served, or whatever metric is relevant.

Consider scope. Some projects may be multi-year collaborations on complex policy areas, while others may have quicker turnarounds. You may want a combination of both in your repertoire.

Finally, mayoral priorities must be a consideration. Most mayors and city administrators will have a strategic plan or list of goals. These are a good start for finding projects that have a mandate to innovate and the appropriate resources for implementation.

4.3.3. *Working with Other Departments*

How can the analytics team be an effective partner to others?

By definition, a city's analytics team is a partner to other departments.

This is important to keep in mind because it should dictate the attitude taken when relating to internal stakeholders. After all, if anything goes wrong, it's the department that's directly implementing a service that usually gets blamed, not the analyst. Avoid the negative spotlight. Don't assume a project that is fascinating to you will be of interest to whichever commissioner or project director you want to pitch it to. Understand their priorities and avoid politically sensitive subjects when you don't have the appropriate cover.

Instead, when approaching potential partners actively listen to their goals and concerns. When you do have an idea to bring up, make sure to frame it in a way that is attuned to their interests: "We would love to sit down and discuss how data could help you handle this problem or achieve this target."

Another good practice in early exploratory meetings is to **provide a variety of ways of helping**. For example, you could:

- Provide advisory functions upon request
- Help research a particularly thorny question through data
- Offer training to people in the department if they have to use data or a specific software tool in their day-to-day functions
- Assist in developing a procurement strategy for tools that would best serve their need and allow for better analysis

Finally, remind your partners that you will be in the background and will do whatever it takes to make sure they get recognized for the success of the project.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

4.3.4. Data Stewardship

How do you think about data management?

Eventually, the team will need to start thinking beyond building effective collaborations that lead to fruitful projects and toward setting the right data infrastructure, or how data is managed, organized, and governed. This is known as data stewardship, and it may be vital to the long-term success of your analytics team.

Some cities, like Boston, have installed centralized data warehouses built in collaboration with contractors and maintained internally (see case study below). Such a tool may or may not be the best alternative for your city; what's most relevant is the concerted and iterative effort to rethink the city's data infrastructure, and how that can be connected to the mission of the analytics team.

Case Study: Boston's implementation of a centralized data warehouse¹⁵

The City of Boston has implemented its own centralized data warehouse, which serves as a repository for different databases that can be used throughout several departments. It was built over a period of three years with a contractor hired through a competitive bidding process. The project started small, encompassing a handful of databases, automating the loading of data into the system as much as possible, and expanding from there. Today, more than 30 departments have their data up and running.

Boston decided to make this investment in data infrastructure primarily for two reasons. First, to create a central repository (a single source of "truth") and avoid the duplication of data across several departments. Second, to shift the work of data analysts from tracking down and cleaning data to adding value to the data by, for example, understanding the operational implications behind each project or ensuring the robustness of the analysis.

Boston's warehouse has been instrumental in implementing one of the city's priorities: its Vision Zero for eliminating fatal and serious car crashes,¹⁶ which combines data sets from multiple stakeholders.

A longer write-up of this case can be found in the annex.

4.3.5. Pushing Projects to Production

How can you turn insights into value?

In whatever realm the analytics team in your city may hope to make a difference, it should ultimately strive to not only create insightful analysis, but also to develop products or tools that are useful for other employees of the city and that (either incrementally

.....

¹⁵ Interview with Maria Borisova, Data Engineering Manager at the City of Boston (February 2020).

¹⁶ <https://www.boston.gov/transportation/vision-zero>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

or radically) change their day-to-day operations. This requires more resources and a larger team, but it is the key to delivering impact and innovation.

In trying to address this challenge, the boss at Transport for London's data team (see case study below) uses philosophies that are not always common in local government: agile development—a way of organizing work around iteration and quick prototypes—and the introduction of product management to oversee the continuous and iterative improvement of the applications that generate business value, as opposed to project management where projects have start and end dates.

Case Study: Transport for London and leveraging data products¹⁷

As the chief data officer of Transport for London (TfL), Lauren Sager Weinstein heads a team of 70 people, including data scientists, product managers, software developers, and data architects. The team is relatively new in the organization, and part of its mission is to centralize the creation of tools that use the vast amount of data generated by London's transportation network—from traffic information to traffic-signals data to costumer data—while preventing the creep of siloed data tools that traditionally didn't interact with one another.

TfL's analytics team is particularly concerned in creating organizational change, since if there is no connection to actual operations that will be affected, then there is no real business value. To do this, they have adopted strategies such as:

- A product management mindset, where product managers continue to work closely with operational experts who use the results of an analytics project
- An agile methodology to their projects, which includes the creation of minimal viable products with defined outcome metrics
- A commitment to transparency, privacy, and clear communication of their work and its value to the general public

A longer write-up of this case can be found in the annex.

4.3.6. Branding and Communications

How should you communicate the analytic team efforts?

You need to be proactive about communicating the work of the team.

One reason is that an effective branding strategy will help you with the other challenges we have pointed out: attracting the right talent to the team, building meaningful partnerships with other departments, and securing appropriate resources.

.....
¹⁷ Interview with Laura Sager Weinstein, Chief Data Officer at Transport for London (March 2020).

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Another reason is that there may be (often well-founded) skepticism and challenges to using methods such as predictive analytics in government. If these fears are not addressed they may paralyze the work of the team.

The most appropriate strategy for the analytics team will likely depend on a combination of the organizational choices and the current state of the conversation. It will require different approaches in different situations; depending, for example, on whether you are working on concrete operational issues or pursuing one of the mayor's priorities, which is likely to have a marketing strategy of its own.

In any case, be prepared to explain in simple terms what you are trying to accomplish, while also having considered the implications of your work (see Section 4.4.2, "The pitfalls of analytics").

As much as possible, make your data, code, and insights publicly available. Academic institutions will probably be willing to partner up and further analyze your results, which will help you spread the information to an even wider audience as well as validate your approach.

4.4. Important Considerations

In this section we include other key concepts that will be vital for the analytics team. Although we offer only a brief explanation for each of these challenging concepts, we believe that the people behind analytics in government should keep open discussions surrounding how to remake government from a platform perspective, the importance of open data and sharing, the ethical considerations behind this work, and, above all, the understanding that when poorly managed, these projects can actively cause harm.

4.4.1. Open Data

How can you use open data requirements as an advantage?

Across many jurisdictions, there is an increasing legal or political mandate for open data policies.¹⁸ In New York City, for example, legislation approved in 2012 demands that all departments upload their data using open data standards. The push has often been led by civil society with the aim of increasing transparency and predicated upon established principles such as accessibility, timeliness, and completeness.¹⁹ Governments may also have an interest in publishing their data as it may lead to civic innovations.

.....
¹⁸ <https://www.govtech.com/data/Are-Governments-Committed-to-Open-Data-Interactive-Map.html>.

¹⁹ <https://project-open-data.cio.gov/principles/>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

The call for open data within city hall, whether as a formal decree or an informal demand, can be a great way to speed up the potential for analytical work by giving you more data sets to play around with and build the data infrastructure of the city. But you should consider some caveats.

- Not all mayors and city managers will be equally receptive (or pressured) to push open data. Therefore, the willingness of departments to comply may vary. Understand the history of your city’s open data efforts, the current landscape, and the applicable laws and regulations to open data, and adapt accordingly.
- Even if there is a clear mandate, you should watch out for potential duplication. If there is a legal requirement, there may be an office separate from the analytics team that has been set up to help departments with open data compliance. Thus, certain departments may have to send around the data multiple times.
- Adding data to an open data portal is costly. It will probably involve an active process of data cleaning to scrub away any personally identifiable information.
- You will have to do some thinking around your users. Most open data efforts assume that there is a standard generic user and they tailor their efforts to whatever is most convenient to the city.

Lessons from NYC: Considering multiple consumers of open data

In reality, you may have different types of users, with different skill sets and different understanding of what data is useful or not. When Amen was in MODA, he commissioned the social impact firm Reboot to do a study that found at least six personas or types of users who were using, or could potentially use, their open data portal—mappers, liaisons, interpreters, explorers, bystanders, and community champions.

Several companies offer prepackaged open data solutions that may be cheaper than building your own portal from scratch, but one thing to consider before committing to one is whether that solution has the right users in mind. Another alternative may be inviting citizens to give input into the city’s open data policies; it’s more expensive but you will reap the benefits of a co-created solution that engages the community.

4.4.2. The Pitfalls of Analytics: Privacy, Security, and Algorithmic Bias

What should you, the team, and city leadership watch out for?

As governments start to ingest ever-growing quantities of data, new pitfalls appear. The more granular, rich, and timely your data, the more it will grow in usefulness for analytics, and the higher the risk of potential breach or misuse. Entirely new literatures have emerged over the past decade surrounding the challenges of privacy, security, and bias

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

in algorithms, and universities have created graduate programs for professionals to specialize in these areas. Here we only offer a glimpse of the general things to consider.

Privacy

There is an inherent tension between transparency and privacy, even when you are careful with personally identifiable information. As more and more information is made available in open data portals, concerns arise that something like data on parcels can be linked to individual behaviors. And even if you attempt to anonymize or de-identify certain data sets, studies show that information can be reidentified with a lot more ease than was once thought.²⁰

It is important to listen, respond to the concerns of citizens, and ensure that protocols are put in place in accordance with national, state, and local laws. Understand too that agencies that produce and control data may have their own concerns. Different departments may have different privacy requirements and if you are trying to bring that information into a centralized system, you should be aware of all the limitations.

Cybersecurity

In recent years the number of cyber attacks targeting state and local governments has grown considerably in the United States and around the world. The role of the chief information security officer (CISO) has been getting more attention.²¹

This presents two relevant considerations for the analytics team. First, it should understand that bringing information together under a centralized data infrastructure may create additional vulnerabilities. In the United States, some data types (such as medical information) have specific standards for security that must be met.

Second, the team needs to understand who has the mandate to deliver cybersecurity, which depends on the city's structure. In some jurisdictions this mandate may not even be realized yet and may need to be created. If there is an IT team that is separate from the analytics team and they have been charged with maintaining data security, that team should be a close partner in helping the analytics team plan the next step for the city's data infrastructure.

.....
20 <https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/>, for example.

21 <https://statescoop.com/tag/chief-information-security-officer-ciso/>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Algorithmic Bias

Many academic and journalistic articles have been published in the past couple of years regarding the fairness of algorithmic decision-making. Some worry that black box algorithms built on faulty data may perpetuate or even accentuate patterns of inequality already in place. Others retort that with the right precautions and oversight, data can lead to a fairer distribution of resources by being quicker to identify those who need them or by predicting where an intervention can have the largest impact.

Some local governments are taking steps. New York City enacted an ordinance to monitor its automated systems,²² and states like Vermont are developing guidelines.²³

While there are no simple solutions at the moment, a good starting place is transparency. Having a scrutable open-source library of every project and auditing any automated decision-making system in the city should help create some trust. One useful resource to consider is the Open Data Institute’s Data Ethics Canvas,²⁴ which encourages analytics teams to be thoughtful about the primary purpose of any project and consider those who may be negatively affected by it.

.....
22 <https://www.jdsupra.com/legalnews/nyc-s-task-force-to-tackle-algorithmic-93703/>.

23 <https://statescoop.com/vermont-releases-first-state-level-ai-report/>.

24 <https://theodi.org/article/data-ethics-canvas/>.

5. SUMMARY: DATA ANALYST, GO OUT THERE AND LISTEN

In the summer of 2015, cooling towers were killing people in New York City. Legionnaires’ disease is a form of bacterial pneumonia spread through the inhalation of infected water vapors, and several sites in the South Bronx saw enough cases in July and August that it became the biggest outbreak of the disease in the city’s history. Public health officials mandated the inspection and disinfection of all cooling towers in the area, but there was no official record of where such towers were located.

MODA was called to help.²⁵ There had been a dozen fatalities and the city was scrambling to identify and begin tracking all the existing towers. Many departments tried to get as much information as possible in a variety of ways, from creating a self-registration portal to dialing building managers to physically inspecting buildings, resulting in disparate data sets with no “ground truth.”

By quickly pulling together all the information gathered, MODA was able to create a single, reliable data set for the inspections to work from as well as predictive models to find the missing towers. An early machine-learning model had enough predictive power to find approximately 90 percent of all cooling towers in New York. This allowed the city to move quickly, stopping the spread of the disease in a matter of weeks.

Of course, not every project is so dramatic. Some projects may seem more mundane. Some may never be implemented. But they are all important. As MODA’s support in the Legionnaires’ disease crisis illustrates, adding more analytical capacity is increasingly important as cities must adapt to the growing complexity of whatever urban challenges they are facing, and ultimately provide value to citizens. Being able to find previously unseen value in your data sets, ensure that city operations are tied to increases in performance, prove the value of new policies with data, and understand the limitations and pitfalls of these approaches so that they are used responsibly will inevitably go from being seen by city leadership as nice-to-have to must-have.

.....
²⁵ MODA documented this project at <https://moda-nyc.github.io/Project-Library/projects/cooling-towers/>.

6. ANNEX

6.1. Case Study: Chicago's CDO Has a Centralized Mandate²⁶

Tom Schenk Jr. became the second Chief Data Officer (CDO) to serve in Chicago, after having played roles in the private, public, and academic sectors and publishing a book on data visualization. His arrival came just after the department had just been moved from the mayor's office to the Department of Information Technology (DoIT), a department with highly qualified staff but a lot of attrition, which had operational responsibility for maintaining and upgrading every single database and major digital platform used by other departments. The appointment gave Schenk dual roles as both CDO and deputy director of IT.

Dual Roles that Converge

Because he had to understand the ins and outs of the systems that were serving the rest of City Hall, these dual responsibilities provided a tactical advantage. Tom and his team optimized their systems while also molding them to be more responsive to the needs of future analytics teams. This reduced the barriers between analysts and IT professionals, and whenever a database was hard to access, Tom could easily get his team to give access and solve the issue.

Being able to inventorize the data was essential for setting up future successes. However, Schenk warns that quick wins are an absolute necessity. "Stuff will get hard and you will need to ask a lot of people [for help]," he says. "An inventory of data will take a lot of time and effort, and both residents and internal stakeholders will not perceive its value until you show some progress and real, tangible results."

As CDO, Schenk undertook a broad variety of projects, from predictive analytics to open data to automation. One of his most notable projects was a predictive algorithm that identified where kids were most likely to suffer from lead poisoning before they were 1 year old.

Selecting Projects as a Centralized Analytics Unit

As a centralized data analytics unit, how to choose which projects to spend time on was a constant question. The team worked closely with the University of Chicago. In conjunction with the Master of Science in Computational Analysis & Public Policy, they created a framework to screen projects and assess the data maturity required

.....
²⁶ Interview with Tom Schenk (October 2019).

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

for meaningful analysis.²⁷ Although the criteria were very explicit, the project intake remained flexible as the team evaluated the availability of data, the potential impact, and the ability for the partnering unit to operationalize the problem, or what they would do after the analysis was conducted.

Effective Projects through Collaboration

Another publicized initiative was a predictive model for food inspection evaluations.²⁸ This model, which aimed to optimize the procedure for inspecting restaurants and resulted in critical violations being found seven days earlier, sprouted as part of a grant awarded by Bloomberg Philanthropies to Chicago as part of the Mayor's Challenge.²⁹ It later became a collaborative partnership between the Chicago Department of Innovation and Technology (DoIT), the Department of Public Health (CDPH), the Civic Consulting Alliance, and Allstate Insurance.

Schenk also spearheaded the creation of a robust open data policy and hosted hackathons to promote the use of city data and get the community engaged.

An example of the difficulties of collaboration came from one project that intended to predict where illegal cigarettes sales occurred. Because of the nature of the issue, the relevant data were collected by both the county and the city, and getting different levels of government (which have different data collection and standardization practices) onboard required significant coordination. Because the project was not able to get to concrete results early on, it eventually got sidetracked and discarded.

6.2. Case Study: Piloting London's Office of Data Analytics³⁰

In 2016 a pilot began for the London Office of Data Analytics (LODA). It was orchestrated by the Greater London Authority; 12 of the 36 London boroughs; Nesta, an innovation NGO; and ASI, a data science firm.

Although it was modeled after New York City's Mayor's Office of Data Analytics, LODA faced an additional layer of complexity because each of London's Boroughs had its own council, government structure, and data sets without any obvious standardization.

.....
²⁷ http://www.datasciencepublicpolicy.org/wp-content/uploads/2018/05/Data_Maturity_Framework_4.28.16.pdf.

²⁸ <https://datasmart.ash.harvard.edu/news/article/open-data-in-chicago-a-comprehensive-history-311>.

²⁹ https://www.chicago.gov/city/en/depts/mayor/press_room/press_releases/2015/july/mayor-emanuel-announces-new-innovative-system-to-prioritize-food.html.

³⁰ Sources: <https://londondatastore-upload.s3.amazonaws.com/LODA%20pilot%20report.pdf>, <https://data.london.gov.uk/loa/>, <https://data.london.gov.uk/city-data-analytics-programme/>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

LODA found its first project by asking for suggestions from the participating boroughs. Then, a workshop session was conducted to collaboratively assess the projects based on:

- Money saving potential
- Availability of data
- The ability to produce insights and delivering results within two months
- The ability to solve the problem without personal data sets

Data to Tackle Common Challenges

After a multistage project selection process, the pilot focused on one use case: leveraging predictive analytics to identify multiple occupations in houses that did not have the appropriate licenses. By combining data sets from multiple sources, the team sought to point inspectors towards infractors.

One of the main questions the pilot sought to answer was whether this methodology would be easily scalable to all boroughs of the city or to other policy areas. As an ultimate goal, this provided the opportunity to intelligently design shared services and coordinate the action of different teams.

Data scientists from an external consultancy were brought onboard, as the data analytical capacity inside the boroughs was extremely limited and their analysts could not commit full time to this project. Internal and external analysts worked together to identify relatable data sets from building inspections, noise complaints, tax bands, etc.

Learning from Failure

The LODA pilot failed to produce meaningful results or a scalable methodology.

One of the main barriers was that the boroughs collected widely varying data, and when they did collect the same data, they used different formats. This forced the team to create individual models for each borough. Much time had to be spent processing, cleaning, and merging the data. Moreover, the data sets weren't properly geotagged; that is, there wasn't a unique identifier for each property that allowed the merging of data.

From a data science standpoint, it was hard to carry out a predictive model because the variable of interest was only half-labeled; the team knew for certain when some houses fell in the "unlicensed multiple occupation" category, but didn't know for certain when a house was definitely not in that category. This made assessing the accuracy of the model difficult.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

Other challenges cited by the pilot group included:

- **Data quality:** significant effort was devoted to cleaning; inability to geomatch certain data sets
- **Data availability:** private rental data were missing or hard to access
- **Data warehousing:** some boroughs did not have centralized business intelligence units or data warehouses and the data had to be pulled individually from different sections of the organization
- **Rarity of the predicted variable:** The variable of interest was too rare in certain boroughs, which made the construction of a predictive model hard
- **Lack of capacity:** Lack of available in-house expertise in the boroughs to work in both the data analytical portion and the implementation that should have followed

After the pilot, an information sharing protocol was signed by 12 of the boroughs and Mayor Sadiq Khan announced the creation of a City Data Analytics Programme to build on the connections made possible by the data partnership between boroughs.

6.3. Case Study: Boston's Implementation of a Centralized Data Warehouse³¹

The City of Boston has implemented its own centralized data warehouse, which serves as a repository for different databases that can be used throughout several departments. It was built over a period of three years with a contractor hired through a competitive bidding process. The project started small, encompassing a handful of databases, automating the loading of data into the system as much as possible, and expanding from there. Today, more than 30 departments have their data up and running.

Investing in Data Infrastructure

Why is Boston investing in its data infrastructure? On one hand, the city avoids conflicting reports of data by establishing a single repository of reliable information, which creates trust. Before establishing a centralized warehouse, several versions of the same data may have existed in different departments. For example, if the 311 phone line (which in many cities is the main aggregator for citizen requests and complaints) gets reports about potholes in the streets, it may keep a separate list and then pass it on to the Department of Transportation, which is in charge of filling the potholes. The 311 division and DOT may keep separate data sets, perhaps with different attributes. (Has the complaint been inspected? Has the citizen been contacted when his case

.....

³¹ Interview with Maria Borisova, Data Engineering Manager at the City of Boston (February 2020).

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

is closed?) If you wanted to understand the entirety of the pothole operational performance, you would have to track down several different data sets held by different people in different departments, perhaps even in different formats.

This helps clarify a second advantage of a centralized warehouse: as you move toward enacting advanced analytics—whether by building dashboards, running predictive models, etc.—a centralized model will save your analysts will considerable time and effort and their results will be more reliable. Because they don’t have to spend tracking down and cleaning the data, steps that could have taken weeks or months turn into days, and the analytics team can spend its time on activities that really add value, such as working with partners to understand the operational implications behind the data or ensuring the robustness of the analysis.

Boston’s warehouse has been instrumental in implementing one of the city’s priorities: its Vision Zero for eliminating fatal and serious car crashes.³² This plan, which involves several departments (transportation, public works, police, and more), is built upon reporting, dashboards, and geospatial visualizations that rely upon data that is also collected by several stakeholders. By leveraging the data warehouse, the analytics team at the Department of Innovation and Technology was able to quickly and frequently provide information regarding traffic patterns, interventions at streets, reported incidents, and more.

Maria Borisova, one of the city’s software engineers who has been overseeing the warehouse implementation from its inception tells us that the technical details are relevant but not that hard to figure out. It is often working with other partners that requires thoughtfulness. Plenty of times Borisova’s counterparts at other city departments may have concerns about centralizing their data, worrying about the accessibility, reliability, and safety of the new system. Patience, starting small and building iteratively, and showing value to your counterparts is vital, she says.

6.4. Case Study: Transport for London and Leveraging Data Products³³

As the chief data officer of Transport for London (TfL), Lauren Sager Weinstein heads a team of 70 people, including data scientists, product managers, software developers and data architects. The team is relatively new in the organization, and part of its mission is to centralize the creation of tools that use the vast amount of data generated by London’s transportation network—from traffic information to traffic-signal data to

.....
³² <https://www.boston.gov/transportation/vision-zero>.

³³ Interview with Laura Sager Weinstein, Chief Data Officer at Transport for London (March 2020).

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

customer's data—while preventing the creep of siloed data tools that didn't interact with each other.

Creating Real Business Value through Analytics

The data science team at TfL seeks not only to understand data but also to create meaningful products that add business value to the organization. This requires an understanding of TfL's strategic priorities,³⁴ such as expanding bus service to outer London and reducing carbon emissions, as well as of the operational complexity of the network. While keeping constant communication with other divisions of TfL, a data scientist may find an insight that could be used to improve more processes. After a proof of concept to test its usefulness, a team of developers may build a software tool around it and a dashboard to measure its outcomes. One of the several product managers may be in charge of ensuring that the tool continues to fulfill its goal and hit its targets.

To accomplish this, the team must work with an agile methodology, building minimal viable products with defined outcome metrics. It is essential that any partner(s) for a given project should have a clear articulation of what analysis they need and what it will be used for. Without a connection to actual operations that will be affected, there is no real business value. For example, to execute the city's Vision Zero,³⁵ which seeks to eradicate road deaths by 2030, the transit division wanted to understand where bus speeding was most frequent to then take the right preventive measures.

Another of the team's recent undertakings has been a pilot to test whether WiFi usability data (collected at tube stations) could help the organization's understanding of traffic patterns and help create measures to avoid congestion, improve operations, and prioritize investments. With data coming from over 500 million connection requests, the team prototyped a series of solutions, including the display of approximate train congestion for passengers waiting on the station and others. By using an agile approach, TfL has been able to test both the technical and business value feasibility, and can now begin to consider moving to a production version of some of the ideas that were tested.



WiFi connection data could help us improve customer information for journey planning and avoiding congestion

³⁴ <http://content.tfl.gov.uk/tfl-business-plan-2019-24.pdf>.

³⁵ <https://tfl.gov.uk/corporate/safety-and-security/road-safety/vision-zero-for-london>.

A Navigational Tool for Understanding Data Analytics in Local Government, Confronting Trade-Offs, and Implementing Effectively

It is also worth noting TfL’s commitment to transparency—the results of pilots are published online³⁶—and clear policies regarding privacy and the use of personal information.³⁷

.....
³⁶ <http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf>.

³⁷ <https://tfl.gov.uk/corporate/privacy-and-cookies/privacy-and-data-protection-policy>.



A publication of the

Ash Center for Democratic Governance and Innovation
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138

617-495-0557
www.ash.harvard.edu