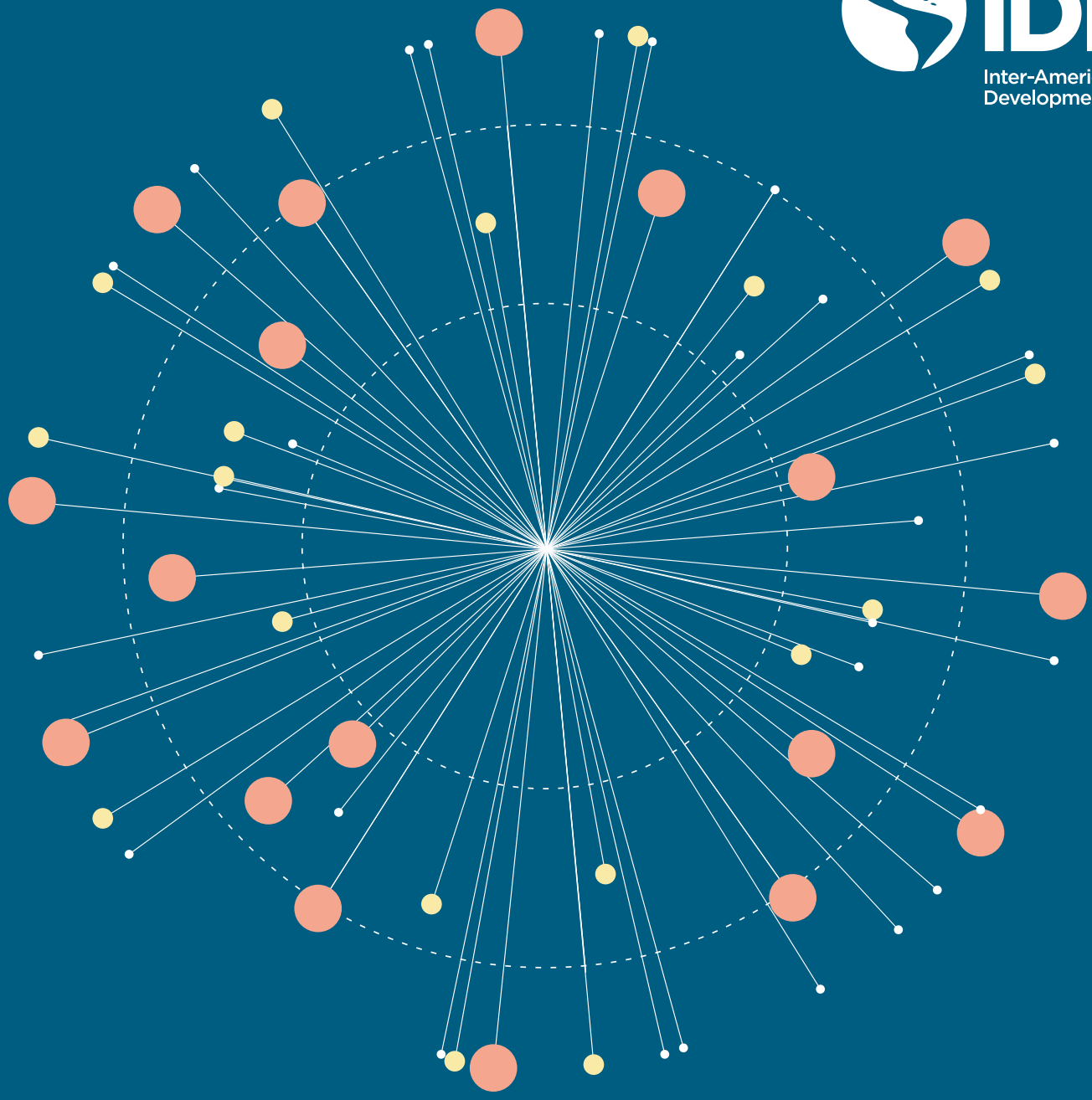


From Information to Actionable Intelligence:

Adapting Governments to Data Analytics

Prepared for the Inter-American Development Bank by:

Mike Bracken, Andrew Greenway, and Angeles Kenny

Copyright c 2019 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNICITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that the link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

Banco Interamericano de Desarrollo
1300 New York Avenue, N.W.
Washington, D.C. 20577
www.iadb.org

The Institutions for Development Sector (IFD) and Social Sector were responsible for the production of this publication.
English editing: Eva Fernandez lopez de Turiso
Design: Gaston Cleiman

From Information to Actionable Intelligence:

Adapting Governments to Data Analytics

Mike Bracken, Andrew Greenway, and Angeles Kenny

IDB
Inter-American Development Bank

Index

Acronyms

CONOCER	National Council for Standardization and Certification of Labor Competencies
HDX	Humanitarian Data Exchange
ICT	Information and Communications Technology
MODA	Mayor's Office of Data Analytics
OCDE	Organization for Economic Cooperation and Development
OSM	Open Street Map
PLJ	Pulse Lab Jakarta
UN	United Nations

Introduction

Much has been written about governments around the world undergoing a data revolution. Open data, big data, massive data, inclusive data—a new language has grown up to describe the changes happening on the quantitative side of digital transformation. More and more countries are launching ambitious programs aimed at improving their use of data. These programs are beginning to tackle governments' ability to recognize, structure, and exploit this valuable resource, and, by opening public data in more accessible and user-friendly ways, create the means for others to generate value from it too. Many governments are starting to recognize that the state of their data infrastructure—the reliable sources of information that all administrations need to build internet-era public services and extract useful policy insights—is not sufficiently reliable. For advanced economies, relatively simple things such as comprehensive lists of addresses are often duplicated or incomplete. For less developed nations, essential, canonical resources are often not in place at all.

The process of applying internet-era data analytics to public policy challenges has so far been one of evolution rather than revolution. Data is widely seen as a necessary tool for helping to solve real political problems, but there is also a growing awareness that even with the right in-

frastructure in place, simply having a lot of data and the tools to manipulate it is insufficient. Public bodies are having to reconsider the skills and ways of working that are needed to support the effective use of data and analysis to improve their programs and services. Realizing the benefits of data-driven government often requires revolutionary thinking in terms of what governments do and how they go about doing it. Making the most of data ultimately demands organizational as well as technical changes. Without doing both, it is very difficult, if not impossible, to shift from policy-driven data-making to data-driven policymaking.

Some governments around the world have already taken steps toward data-driven policymaking. Several have put initiatives in place to make the most of their data to improve infrastructure, internal processes, or public services. As one would expect in an evolving field, there is variation in the strategies and tactics adopted by these initiatives. The emphasis of the projects has varied, with some being led more by their technology choices (e.g., Dublin, Chicago, Singapore) and others by opening up data (e.g., New York City).

However, several common trends are beginning to emerge among the efforts currently having the most significant impact on public life. These projects' team members tend to have multidisciplinary skill sets, working across organizational silos rather than creating new ones. They also focus on ensuring their choices on technology and data are made to serve user needs.

Put to its best use, data analysis is not an end in itself or another form of internal corporate reporting. It offers a powerful set of insights used to amend or redirect public services on the frontline, making them work better for the citizens and businesses that rely upon them. Strengthened data analysis has helped governments become more effective at delivering on their policy objectives and running public services, which ultimately improves the lives of their citizens.

While many data analytics initiatives share certain qualities, this review of best practices is intended to show that successful data analytics teams are not all alike. In fact, they can be set up with quite different governance structures, as appropriate to the institutional context of the government they work in. Where a data analytics team sits in the organizational structure, who it reports to, what its mandate and scope is, who its external partners and stakeholders are, and what skills make up its team can all differ. Yet, there are sufficient similarities in approach to loosely classify the governance models into three categories:

- **Centralized**, where a data analytics team is set within the center of government, under a prime minister, president, or mayor's office, and given a mandate to deliver.

- **Federated**, where data analytics capacities are widely distributed across several agencies

- **Ecosystem**, where there is a coordinating function for data science in government, but it functions primarily to convene partnerships with private institutions and other stakeholders responsible for the data-analysis deliverables

This report presents a case study of New York City's centralized model, Singapore's federated model, and Mexico's ecosystem model. In addition, Box 1 examines Ireland's version of the ecosystem model, and Box 3 looks at Indonesia's hybrid approach to data analytics.

There is no single right way to create an effective data analytics team in government. That said, there are clearly certain characteristics to embrace and some to avoid. We have summarized a short set of recommendations we believe should be considered by all governments seeking to develop or enhance their data analytics expertise.

Mini Case:

Ireland's Ecosystem Governance Model Variation

Given the breadth of the three categories of models, variations are common, and hybrids are inevitable. Ireland offers another example of an ecosystem model. Rather than funding an internal, centralized team to coordinate an ecosystem, Ireland has instead provided central government funding directly to public-private partnerships.

The Irish government has identified big data as an area where the country has distinct advantages in developing private partnerships. It has decided to invest in data analytics capabilities by funding research centers such as the Insight Centre for Data Analytics. This center is in a unique position at the interface of academia and industry, and it undertakes high-impact research in data analytics that has significant benefits for individuals, industry, and society by enabling better decision-making. The Smart Dublin initiative has built yet another type of ecosystem, in which the four Dublin local authorities formed a network to engage with smart technology providers, researchers, and citizens to solve challenges and improve city life.

It is too early to determine how successful this variant of the ecosystem model will be, but it points to alternative routes for funding the development of data analytics capability.

Building a Data Analytics Team: Recommendations and Choices

There are six recommendations for any government looking to set up a successful data analytics team. These should apply equally for almost any country and context, though they will obviously be more challenging to implement in some places than others. Nonetheless, we believe it is unlikely that any government could create a successful, sustainable data analytics team capable of delivering significant public value without doing these six things.

Recommendations

1. Recruit an expert, multidisciplinary team. Creating data-driven solutions for policy issues requires a mix of people with new skills and approaches as well as receptive leadership that grants the team permission to work in a different way. A data analytics team needs in-house developers, social scientists, and designers working together. The developers need to be capable of writing code in modern programming languages, the social scientists need to frame public policy questions, and the designers need to be capable of presenting information in accurate and compelling ways. Flexible and dynamic human resource management is crucial to recruiting and retaining people with these skills and effectively structuring them into teams. To overcome budgetary restrictions, governments may decide to redirect resources from their existing, traditional statistical functions into more modern data analytics capabilities.

A government should not try to pursue data analytics projects with high-stakes results unless it is prepared to invest in at least one team of five or six people who can actually deliver and/or manage partnerships with collaborators outside government. An under-resourced or underskilled attempt will lead to disillusionment and lost momentum before the approach has had a proper chance to demonstrate its impact.

2. Demonstrate value quickly. Successful data analytics teams have focused on delivering visible and significant answers to recognized problems, often in response to a specific public crisis or emergency. Issues with broad political support and high public impact, such as emergency and crisis management, provide the ideal conditions for building an empowered, high-quality team that can demonstrate the potential of data-driven policies.

Specifically, we would urge governments to publish early, impactful examples of data use, such as mapping visualizations that overlay multiple data sources to support emergency or operation responses (as these are often relatively simple

3. Secure political sponsorship. Data analytics teams need visible, senior political backing—ideally from the very top of government—to succeed. Teams often require hard powers to overcome institutional blockages to effective data analysis (such as data sharing and establishing consistent identifiers for particular data fields), drive cross-agency cooperation, and challenge silo mentality. A political sponsor is essential to acquiring these powers and giving the team a platform with sufficient public exposure to make a name for itself.

Specifically, governments should ensure there is support from the executive's central office and finance ministers before setting up a data analytics team.

4. Establish a clear mandate. Once the importance of data-driven policies has been established, through a tangible demonstration of a data analysis team's value and a publicly visible political sponsorship, institutional incentives will need to be aligned so that the team can sustain its influence over the expanding scope of government activity. Depending on the political system, either formal laws, direct-decree commands, or firm organizational guidelines can establish these norms for effective cross-government work. Without some form of formal mandate that goes beyond political sponsorship, data analytics teams struggle to maintain their effectiveness over time and under different administrations.

5. Answer a clear need. All data analytics teams must be able to balance their workload so that they benefit citizens and businesses with real-world impacts while assisting their colleagues with achieving internally focused objectives, such as meeting shared goals and incentives. Additionally, if teams are not operating in response to a clear need, there is a risk they will be perceived as uniquely able to work on whatever they want and therefore separate from others in government. Another risk to teams' credibility comes from allowing a team to start with solutions in mind rather than with a government's desired outcomes for its citizens. Teams taking a purely technology-led approach, such as focusing on buying or building dashboards or designing data-driven apps or other tools without considering the needs to be met risks derailing the project.

Specifically, governments should consider mandating certain activities across the public sector (such as provisions for opening some types of government data) and giving specific ministers or officials formal accountability for some key datasets (e.g., land or business registries) to ensure that their quality and integrity is maintained.

A data analytics team should, therefore, prioritize projects that respond to a clearly stated need that has been pinpointed by another team within the government, rather than only investigating hypotheses that are gathered from the data analytics team itself. Countries have adopted different models for identifying and structuring the demand for solutions based on data analytics, such as through an open call for ideas (Mexico) or an internal "consultancy-style" arrangement where the team's services are paid for (Singapore). Regardless of the procedure used to identify the need, the team must be disciplined in evaluating the potential of the proposed projects and prioritizing them according to public impact.

If explicit demand from internal stakeholders for further work from an analytics team dries up, a government should not throw good money after bad to keep the team operating.

6. Build data literacy across the public sector. While data analytics teams often act as an accelerator and a center of excellence, their influence is significantly greater if other officials are intelligent consumers of data analysis. Capacity-building support and technical guidance often go together with increased institutional buy-in for and awareness of the potential of data analytics across the public sector. Academic institutions and other organizations are often able and willing to collaborate with governments to achieve this goal.

Governments should roll out basic data literacy training, focusing on how analysis can help policy and operational experts build a stronger evidence base for making decisions (rather than teaching them how to do the analysis themselves). This is best done when data analytics projects are delivered, as a means of showing the impact of the work as opposed to learning about the benefits in the abstract.

Tradeoffs in Approach to Data Analysis

Alongside the recommendations, we believe governments considering new data analytics capabilities have four important choices to make based on the political, institutional, and economic conditions of the government. These choices will inform the scope of the analytics functions. On each of these four points, it is important is that the government make a considered, conscientious decision.

1. Change underlying processes while avoiding changes to frontline work or create new data science tools for use in frontline operations. In some scenarios—particularly for governments where there is a high degree of change fatigue or where making changes to day-to-day business is heavily constrained by hard rules—it is a better strategy to deliberately avoid changing the work of frontline staff, as in the case of New York City. Instead, focus on internal changes further up the strategic chain of command. The initial cost of implementing training and new processes at the frontline can be high and time-consuming. Directly interfering with processes may also be perceived as a political nonstarter in large and complicated organizations, given the potential disruption, controversy, and resistance from powerful stakeholder groups.

   Where there are fewer emotional, logistical, or legal barriers to change, data analytics teams have focused their efforts on creating tailored front-end products for staff that do not know how to code, putting data science tools into their hands. When accompanied by adequate training and preparation, these tools have enabled them to hugely increase the government's collective ability to perform independent analysis and enhance understanding about citizens' needs. A consideration in this choice is the political risk associated with rapid, disruptive change to frontline services. Dealing with entrenched union resistance is one example of where the potential disruption may be seen as outweighing the potential value.

2. Use extremely simple technologies or radically new approaches. Solving complex problems that governments have unsuccessfully grappled with for years sometimes demands a mixture of creativity, new quantitative techniques, and experimental technology to find possible solutions. There are emergent-use cases of governments applying machine learning techniques, such as in one of Belgium's regional public employment services, the VDAB, which uses machine learning to improve job matching by providing better and faster connections between job seekers and available jobs.

   Yet the experience of successful data analytics units shows that moving straight to advanced technology is not always necessary. Seemingly intractable issues have been solved by making use of the existing systems, networks, and databases, or even

by running different combinations of data in simple spreadsheets. Applying only simple, widely available technologies has also helped to reduce or eliminate financial barriers to data sharing. A primary consideration in this choice is the degree of technological maturity in the government, and another is budgetary constraints. In governments with limited budget and experience in applying sophisticated digital technology to public services, starting simply with a relatively low-tech approach is less risky and more likely to yield tangible benefits faster.

3. Establish partnerships to deliver collaborative products or build a comprehensive in-house capability. In governments where it has been possible to do so, data analytics teams have tended to prefer hiring in-house specialists. Using full-time employees can help to speed up communication, foster closer working relationships, and create opportunities for sustained collaboration between different branches of government. However, this option is not always available. In such cases, governments have found it more beneficial to establish partnerships between public or private institutions and rely on these collaborative efforts to produce data-driven solutions to public issues. A primary consideration in this choice is the degree of flexibility the government has to hire relatively expensive specialist skills and the general availability of those skills in the local market. If the government's hiring power is limited and local skills are more readily available in academia or the private sector, some form of partnership between a core in-house team and nongovernmental organizations may be a more pragmatic answer to capability building.

4. 4. Demand good quality data or use what is available. Some data analytics teams set strict data quality standards for all agencies as an unconditional requirement for working together. If the available data is not in the right format, the data scientist can add only limited value. Other teams have taken a more pragmatic approach in this situation. This is particularly true for governments where the instincts or incentives for sharing data are relatively weak. In this case, a data analytics team insisting agencies share their data, even when the provided data is not perfect, at least provides a starting point toward data-driven policies. A primary consideration in this choice is the need to determine which battle it is appropriate for the data analytics team to engage in at a given time. If a government's data analytics operations are relatively immature, the establishment of norms for collaboration and data sharing is arguably more important than demanding data purity. In more data-aware governments, there should be a bias toward achieving quality and sustainability, as well as the speed of delivery.

Case 1: **Centralized Model, New York City**

The Mayor's Office of Data Analytics

Photograph from: MODA

Highlights

Long recognized as a leading public institution in data collection, New York City's government has a well-established central intelligence center with a mandate to apply data analytics to promote data-driven public policies. The Mayor's Office of Data Analytics (MODA) was set up in 2013 with an explicit central mandate from the executive office to work across the city's agencies.

MODA's success has been built on the principle of starting simple and small. The team began life as a small team of half a dozen young analysts called "the kids." Initially, it worked with tools no more complicated than spreadsheets and desktop computers. In order to minimize organizational friction, the team deliberately focused its efforts toward changes that avoided making direct interventions in the day-to-day work of frontline staff. While that seems counterintuitive in terms of reform, streamlining existing processes and systems rather than opting for wholesale disruption proved extremely powerful for winning credibility and support.

In exchange for access to MODA's data and expertise—a latent demand that grew as the unit's reputation increased—all agencies needed to first share their data

publicly. This created a clear incentive for encouraging openness across government.

The New York City model has been described as centralized. This is because it is built on a central team that leads the delivery of data analytics itself, using levers to drive impact at scale across other parts of government and sponsored by a single strong political leader.

What Was Done

For many years until the early 2000s, New York City's agencies routinely collected rich data on buildings, streets, infrastructure, businesses (including tax records), building permits, crime-related data, noise, and other citizen's complaints to measure government operations. Much of this was gathered through the 311 hotline, app, or website. However, most of this information was stored in agency silos, impeding the city's collective ability to aggregate, analyze, and synthesize. While this was clearly limiting the value the data could have, the government also lacked any compelling argument to disrupt this institutional setup.

That argument arrived in 2009, with a financial crimes task force set up in the wake of the 2008 economic meltdown. With little in the way of a job description or a template methodology to work from (this was prior to the formal creation of MODA), the task force team quickly realized it could use the different city agencies' data to better target mortgage fraud investigations by examining past cases. Though the work ultimately did not proceed much further (because banks did not want to prosecute debtors and undermine confidence after the 2008 crisis), the task force established the value of using data analytics to address the city's problems and offered a template for how to deliver it in practice.

After the financial taskforce was disbanded, the analytical team began a new project that looked into how data could complement and strengthen firefighters' natural intuition in identifying dangerous buildings. Data from other departments, such as the age of the building or type of business, was used to create a model that could more accurately predict which buildings were most at risk of having serious fires. By going to the most dangerous buildings first, the New York City Fire Department was able to take early action to reduce the number of days that New Yorkers were at serious risk. The department also used technology to automate the process of reviewing and prioritizing 311 complaints. By offering insights without disruption to daily work, and as a result creating demonstrable and large improvement in their services, the data team won the support of the department, frontline staff, and citizens.

Given the increased recognition of the value of the central data analytics function, the government formalized the unit by executive order by creating MODA. As the city's central intelligence center, MODA's objective was to aggregate and analyze data from across city agencies, allowing policymakers to prioritize risk more strategically, deliver services more efficiently, and enforce laws more effectively. Analytical tools would thereafter be applied to deliver better policies for crime, public safety, and quality of life issues.

Since it was set up in 2013, MODA's work has ranged from answering relatively quick data queries to pursuing longer-term strategic initiatives to increase data-driven decision-making in the New York City government. These initiatives fall into five categories: operational support, citywide sharing, disaster response and resiliency, economic development, and open data.

Box 2. The Mayor's Office of Data Analytics' Five Strategic Initiatives

Supporting operations. Work with agencies, including the city's police, sanitation, and parks and recreation departments to analyze the most effective allocation of city resources.

Citywide data sharing. Through partnerships with different government agencies, MODA built the data-sharing platform DataBridge, which merges data inputs from several sources and overlays geographic information to allow the city to perform cross-agency analysis. MODA also serves as a point of contact for external partners contributing to or using city data or developing data-sharing protocols and standards.

Disaster response and resiliency. As part of the city's response to Hurricane Sandy, MODA integrated data from city agencies, United States National Guard surveys of affected residents, and daily outages from the Con Edison power company to allocate disaster response resources and make sure that vulnerable people received attention.

Economic development. MODA works closely with other agencies on the New York City Economic Development Corporation, the New Business Acceleration Team, and NYC Digital to support small-business entrepreneurs. For example, MODA assessed the time required to open new food service establishments and reduced it by an average of 45 days. The office also created the New York Business Atlas for entrepreneurs that contains detailed information about economic activity, demographics, foot traffic, and other key business metrics that entrepreneurs can use to research potential business sites.

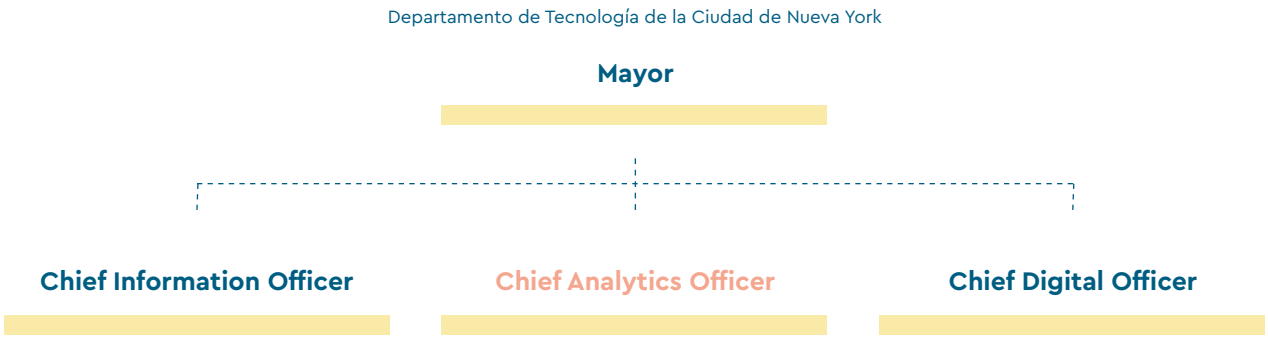
Datos abiertos. MODA also works with the Department of Information Technology and Telecommunications to implement the open data law, whereby all city agencies had to publish their data on the NYC Open Data portal by the end of 2018. Data standards on how agencies should manage or present their data are described in the Open Data Policy and Technical Standards Manual.

How It Was Done

One of the most important drivers behind MODA's impact is the political support from the city's executive offices, giving MODA authority and strong, politically connected leadership. The office became official in April 2013 when Mayor Michael Bloomberg signed Executive Order 306. This established a chief analytics officer in city hall for the first time, as a senior role reporting directly to the mayor.

MODA works closely with the deputy mayor for Operations, Economic Development, and the commissioner of the Department of Information Technology and Telecommunications. Along with the chief information officer and the chief digital officer, the chief analytics officer serves on the mayor's technology council, executing city hall's policy of infusing analytical rigor into all aspects of city operations.

Figure 1. New York City Technology Organization



The chief analytics officer reports directly to the Mayor, together with chief information and the chief digital officers. Source: NYC Analytics (2013).

Rather than creating a significant additional layer of bureaucracy, MODA's model is lean and highly efficient. The staff is a small team of no more than 10 analysts who work alongside technical and administrative support staff. At the core of MODA is the team of analysts, who have a mixture of statistical, public policy, and computer science backgrounds. The analysts have deep experience in data analytics research, government technology, and innovation, public service, and leadership development.

One common source of resistance to working closely with data-driven models is that they can suggest or imply the need for dramatic changes to the ways public officials are used to working. This has led other data analytics teams (especially those that are centrally located and therefore distanced from the frontline) to meet resistance from colleagues and taxpayers. Partly in order to head off this potential criticism, MODA focused its initial efforts on using data related to issues that receive universal political and public support, such as preventing fires or stopping rat infestations.

More generally, MODA's basic strategy has been to avoid changing the work of frontline staff who tend to work directly with the public to deliver services, such as building inspectors. The team has largely discounted interventions that directly change the way the front line works. This choice was made partly for practical reasons; implementing the training and new processes needed to establish new frontline working practice was

seen as costly and time-consuming. More importantly, such changes were perceived as a political nonstarter because of the immense organizational difficulty inherent in turning around large, complicated organizations. MODA's pragmatic decision to avoid this kind of organizational battle offers a salutary lesson for other data analytics units—to go where there is little vested interest to resist the benefits of data analytics and to pick one's battles sensibly.

As an example, if MODA's task is to reprioritize building inspections according to a more sophisticated analysis of the underlying risks, the team integrates more detailed analysis into the existing inspection assignment generation system rather than create a new tool. This means that an assignment is already given a priority level by the time it reaches the frontline inspectors, without them having to adjust their own workflow. If MODA's solution is a technological fix that connects two formerly discrete pieces of information in order to deliver a new piece of insight, the office would make sure that new insight is delivered alongside currently reported data, not in a different, detached method. This concept sounds simple—a light footprint means that the solution must be delivered upstream from the front line—but this underplays the effort required to build things that fit well with existing systems. Nonetheless, doing the hard work to make data analytics simple to use is an investment worth making because however tempting it is to default to building entirely new tools, they create a new type of problem adoption. The New York City government elected to change the outcome rather than the frontline process.

Another work practice behind the team's success was the emphasis placed on collaboration with agencies. As a central team, it was important for MODA to avoid any perceptions developing of a them-versus-us culture between the agencies. To avoid this risk, data models are shaped with experience and expertise taken from the frontline agencies, providing user-friendly context to the data analysis. MODA also makes use of the city's existing data, databases, and networks wherever possible. In fact, the team began with no more than spreadsheets and analysis conducted in old desktops and versions of Microsoft Excel of 36,000 rows of data. The main platforms for data sharing (DataBridge and DEEP, the latter used for secure data exchange) were created to allow agencies to keep using their own systems rather than having to install expensive new IT systems to comply. This means that no major technical change is required from organizations wishing to benefit from MODA's techniques. In cases where minor IT changes are necessary to be able to connect to MODA, a further principle of the New York City model is that those changes are paid for from a central budget. This practice was a deliberate choice to help remove any internal financial barriers to data sharing.

AAs an incentive to share and open up organizations' data, any organization that wants to access MODA's data and expertise must first share their own—a process described in MODA's open Github repository. MODA also insists that while the data provided by an organization does not have to be perfect, it must be the entire set. When MODA collects information from agencies, it is asking them to give them access to their legacy IT systems and share all of their information. Agencies are not obliged to say yes, but they usually do for two reasons. First, by participating in the data exchange, they gain access to the information of other agencies. This has a significant practical benefit: they avoid having to pick up the phone and call the IT department of another city agency for a one-off query; instead, they are able to automatically access the information through the data-sharing platform. Second, agencies default to sharing their data because it gives them access to MODA's skills in helping them to solve problems.

As is typical for all governments, New York City's data projects are rarely straightforward or clearly defined. MODA's analysts are experienced in taking ambiguous problems, and often data of varying degrees of quality, and finding a workable solution. Finding those solutions often demands a mixture of creativity, new quantitative techniques, and experimental technology. When it comes to conducting analytical work, MODA is tool, method, and platform agnostic. The team has avoided committing to any single approach or technology, preferring to focus on a common outcome: using data to create a better-functioning city.

In order to increase the government's data analytics capability at scale, MODA has partnered with the Department of Citywide Administrative Services

to create an Analytics 101 course for city government employees. The course covers basic statistical and data management techniques and provides an overview of available data and tools. MODA is also participating in the department's management academy to promote data-driven decision-making and improve data literacy among city managers.

What Was Achieved

Helping New York City agencies improve the delivery of services to citizens. Before MODA, the first 25 percent of building inspections conducted by the fire department typically resulted in 21 percent of the most severe violations being discovered. With MODA models, the percentage of dangerous buildings identified in the first 25 percent of inspections results in more than 70 percent being discovered.

MODA also helped the New York City Department of Environmental Protection crack down on restaurants that were illegally pouring cooking oil into sewers. By comparing restaurants that had not paid for such a service with geospatial data on the sewers, MODA was able to hand the department's inspectors a list of statistically likely suspects. The result was a 95 percent success rate in tracking down the offending restaurants, thereby increasing the stability of physical infrastructure for citizens.

Sharing data with NYC agencies and encouraging best practice in data analysis. The data that MODA collects and uses is also made available to staff working in 40 other city agencies, enabling them to combine it with their own department's data to improve their decision-making. NYC's Open Data portal sets out clear policies and compliance rules to address ethical concerns around linking personal data.

Using analysis to deliver better insight for economic development. By matching datasets from different civic and commercial databases, MODA created the New York Business Atlas and tracked the average amount of time it takes for businesses to complete all steps of the city's regulatory process for opening a new business. It found that companies that took advantage of the atlas's services opened their doors 79 days faster than those that did not, a 36 percent reduction.

Aiding disaster response and recovery. Following the devastation caused by Hurricane Sandy in late October 2012, city hall realized there was no publicly held map listing all of the city's businesses. Consequently, it was extremely challenging for officials to know which businesses were most likely to have suffered from problems such as flooding and were therefore in need of support to get back up and running. MODA brought together records from six different databases to complete the map.

Data Analytics Leader Profile: Michael Flowers

"I think of us as the Get Stuff Done Folks, all we do is take and process massive amounts of information and use it to do things more effectively." —Michael Flowers.

Sometimes initiatives are driven by the person in charge, and his or her background and vision make this person the leader that the project needs to flourish. This was probably the case in New York City.

Before being hired in 2009 by John Feinblatt, Mayor Michael Bloomberg's chief policy adviser, Michael Flowers did not know much about computer code—let alone Bayesian statistics. From 1999 to 2003, he worked at the Manhattan district attorney's office prosecuting homicides and drug crimes. When he left law enforcement, he moved to Washington and joined the power law firm Williams & Connolly, and he later took a job with the Senate Permanent Subcommittee on Investigations.

After a year behind a desk, he decided to leave that job too. Looking for something more meaningful, he thought of helping to rebuild Iraq. The next thing Flowers knew, he was heading into the Green Zone, the secure area for American troops in the center of Baghdad, as part of the legal team for the trial of Saddam Hussein.

Most of his work turned out to be logistical not legal. Flowers was responsible for sending investigators to gravesites in the countryside and transporting witnesses against Hussein to Flower's office—without getting either group blown up by roadside bombs. It came to his attention that military officers were using predictive informational techniques to determine where and when the bombs were likely to explode.

On his return to New York City a few years later, Flowers realized that those methods marked a more powerful way to combat crime than he had ever had at his disposal as a prosecutor. And he found a soulmate in the city's mayor, Michael Bloomberg, who in 2009 put him in charge of MODA's antecedent, the financial crimes task force. The rest is history.

Hiring Chief Data Officers

When it comes to finding the leader for data analysis teams, it is difficult to build and establish the right profile because the right leader will be appropriate to a specific government's context. However, some leadership behaviors appear to be common to successful appointments:

- A servant-leadership style that protects and empowers the team, focusing on removing barriers to their achievements rather than commanding and controlling the team.
- A bias toward disruption and challenging the bureaucratic mindset where it serves as an obstacle to delivering outcomes.
- Experience in technology and data applications as well as in driving organizational and cultural change.
- Experience in working with large institutions (preferably in the public sector) and start-up internet-era organizations.

Case 2: **Federated Model, Singapore**

A Hive for Data Science

Singapore is currently undergoing a shift from a federated model to an increasingly centralized model. This report focuses on the Singaporean model between approximately 2010 and 2017, which is when this country reached its most developed degree of federated approach. Later, it added to the general trend towards more centralized digital and data initiatives practiced by governments in other parts of the world.

Federalized approaches are more common in the private sector and are not without merit, but they tend to be most effective in organizations where there is a high level of data analytics capability across multiple business units and strong leadership backing each of them. Singapore initially enjoyed some success with this approach, but in 2017 it was decided that—in line with its broader push to centralize the government's technology capability—a central data analytics function was more appropriate to the institutional conditions that they had evolved.

Photograph from: GovTech

Highlights

Singapore's approach to applying data science to government has changed through years of experience in developing technology-related projects. Over the course of several IT-led initiatives, many of Singapore's government agencies had built up their own data and statistical analysis capacities that were loosely federated across different bodies. While the government had a lot of capability collectively, there was a strong perception that most policy decision-making was not data driven. This was especially true where policy cut across multiple agencies. The central government technology agency, GovTech, was created to shake up how the government used technology, and it contains a data science division.

GovTech brought together a multidisciplinary team dedicated to delivering public policies informed by data analysis, in part to address some of the weaknesses inherent in federated data analytics models. GovTech's mandate is to find "real-world impact" data science projects to work on, and every three months the agency asks itself the extent to which it is having a real impact on the ground. The agency has also created front-end data science products that allow staff who do not know how to code to use them in their work. GovTech's funding model borrows from the start-up world: rather than being a cost center for government, it generates revenue by charging each department that uses its services.

This model is considered federated because it has several distinct data analytics teams with relatively deep and broadly distributed capability embedded in different parts of government, a preference for using distributed funding models that give departments responsibility for choosing priorities, and a strong culture of collaboration.

What Was Done

Since Singapore began its digital transformation in the early 2000s, the government has developed a series of IT strategies, citizen portals, and online accounts for digital services, as well as digital platforms for procurement, taxes, housing, business, and more. These were spread out among the 100+ agencies that make up the government, with no single organization responsible for coordinating the state's collective activity. By 2010, most agencies had developed data and statistical analysis capacity, so the government had made significant progress in its transformation. However, the complexity of having several overlapping IT systems owned by different departments (often outsourced to a range of technology vendors) made cross-government data sharing and policymaking very difficult. As the technology market evolved, there was a growing recognition that Singapore's federated approach to data and IT could be curtailing opportunities for the government to deliver more responsive, innovative, and smarter digital services.[1]

In November 2014, Singapore's Smart Nation initiative was officially launched, and a new team was given the objective of shaking up how government departments used technology. The GovTech agency was also established as a 2,000-person-strong department responsible for steering the country's nationally significant technology projects. IDA Hive (now GovTech Hive), GovTech headquarters, was opened in 2015 with a start-up- and tech-sector culture and was insulated from all the bureaucracy. It was located in the Sandcrawler, the Singapore headquarters of Lucas Films' Industrial Light and Magic. A data science division was given the responsibility of delivering data-driven policymaking.

While many of the data science projects were strategically chosen, the team also applied some entrepreneurial initiative to win political attention for their work. The impetus came after the September 2016 disruption of Mass Rapid Transit trains on the Circle Line. Trains would stall for seemingly no reason during the rush hour period, apparently due to "intermittent signal interference." When this same problem happened again two months later, a team of GovTech data scientists was caught in the delays and decided to step up. After spending one Saturday tinkering around with the data, the teamdiscovered

1 From May 2017 onward, GovTech reported to a new Smart Nation and Digital Government Group, under the auspices of the prime minister's office. This change was made to enable a more centralised and focused approach to delivering on a digital government in a smart nation, representing a shift away from a federated approach.

a pattern common to both incidents. Further investigations revealed that one of the trains was emitting a signal that was jamming the signaling mechanism of the tracks.

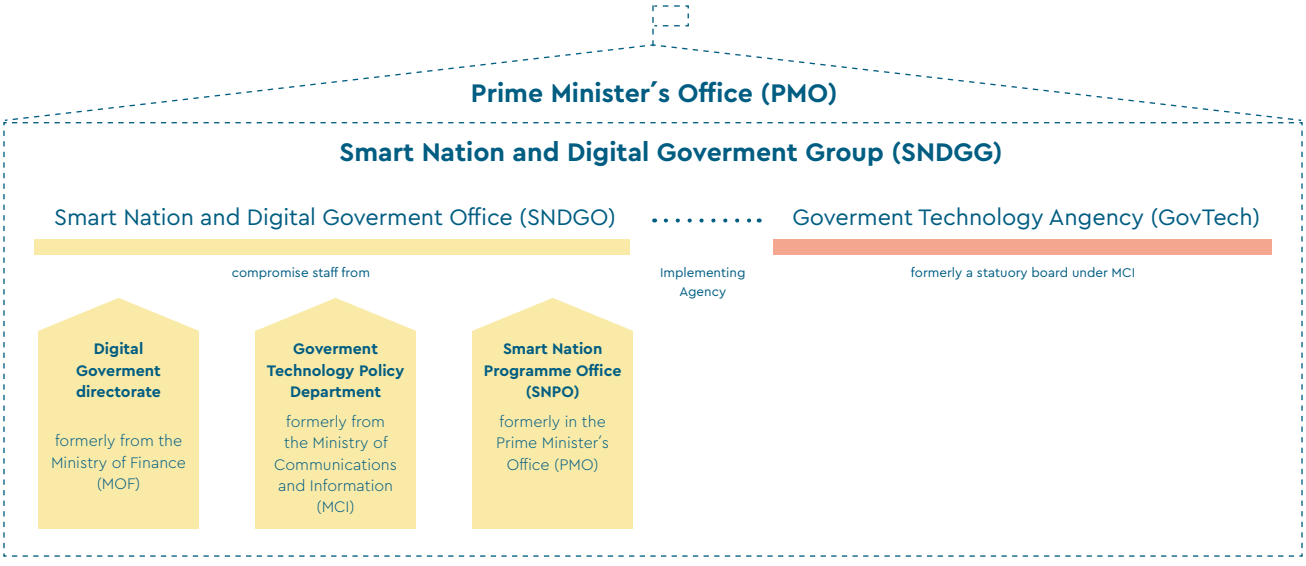
Through a Data.gov.sg blog post, GovTech shared the team's account of how it used data to catch the rogue Circle Line train. The post made national headlines and caught the eye of the prime minister. On his Facebook page, he described the blog post as "a fascinating account, demonstrating close teamwork, sharp analysis, and a never-say-die attitude," adding that "this is how a Smart Nation should use data to solve real-world problems." The Circle Line disruption of 2016 was a perfect example of how an analytics team operating with cross-agency collaboration, plus the right data and skills, could deliver tangible service improvements quickly and at scale. The combination of data analytics and open communications proved very compelling in terms of achieving public and political visibility.

How It Was Done

According to GovTech's vision, three conditions were important to integrating data science into in government: leadership, the right kind of culture, and a public demonstration of success.

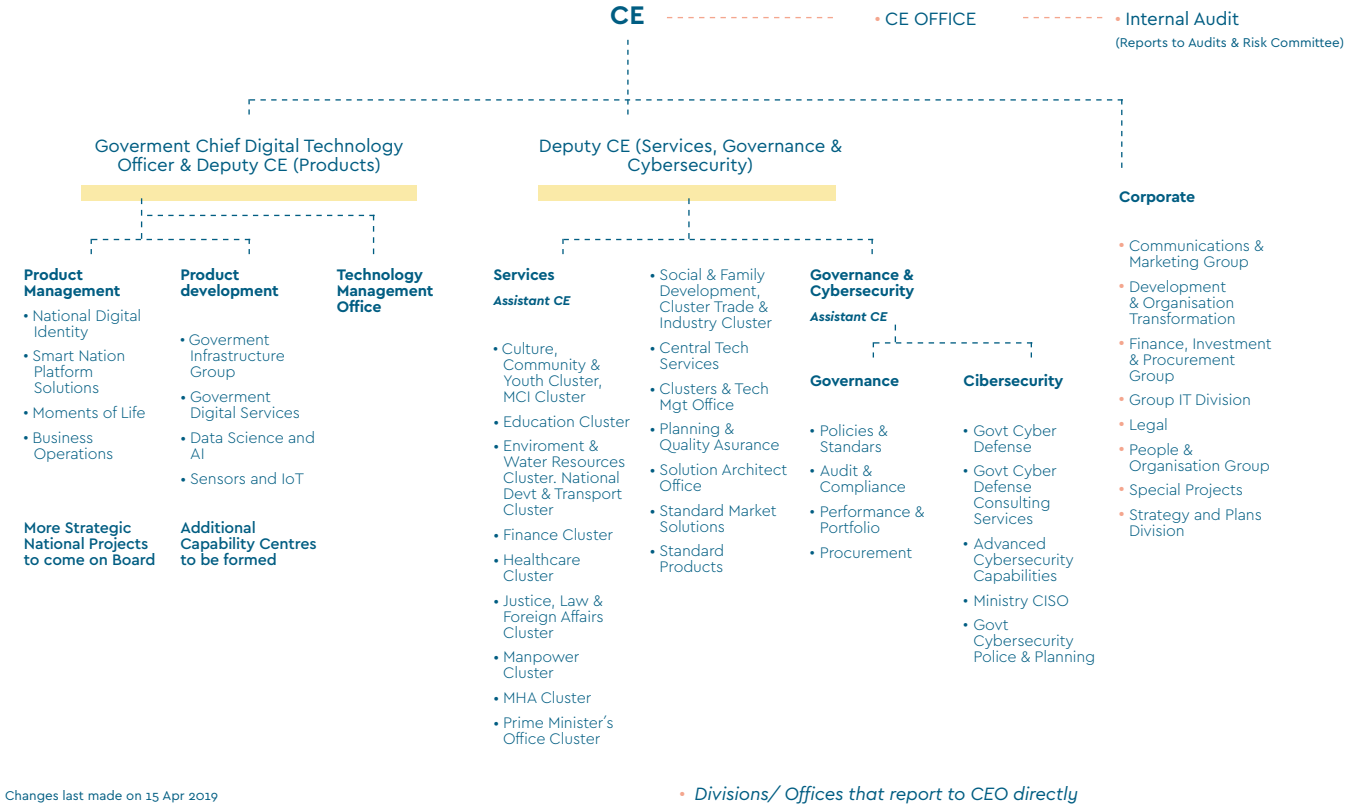
Support came from the top. GovTech won the support of prime minister who was looking to ramp up Singapore's Smart Nation strategy. As part of the Smart Nation and Digital Government Group, GovTech became the implementing agency of the Smart Nation and Digital Government Office. A committee of five ministers was created and chaired by the deputy prime minister himself.

Figure 2. Organisational chart for Smart Nation and Digital Government Group (SNDGG) in the Prime Minister's Office



Source: Prime Minister's Office, Singapore (2017).

Figure 3. GovTech Organization Chart



Source: GovTech

This leadership played a key role in one major factor behind Singapore's federated model's success: the continuous attraction of talented and ambitious staff. The leaders made a conscious decision to hire in-house specialists rather than procure consultants. Full-time employees would speed up communications, foster closer working relationships, and create opportunities for sustained collaboration between different branches of government. But at first, the leaders found out that no one in the government could actually code. Among the 1,000 staff members, only seven individuals had the skills to form a coding team for cutting-edge projects. These people were the pioneering members at GovTech.

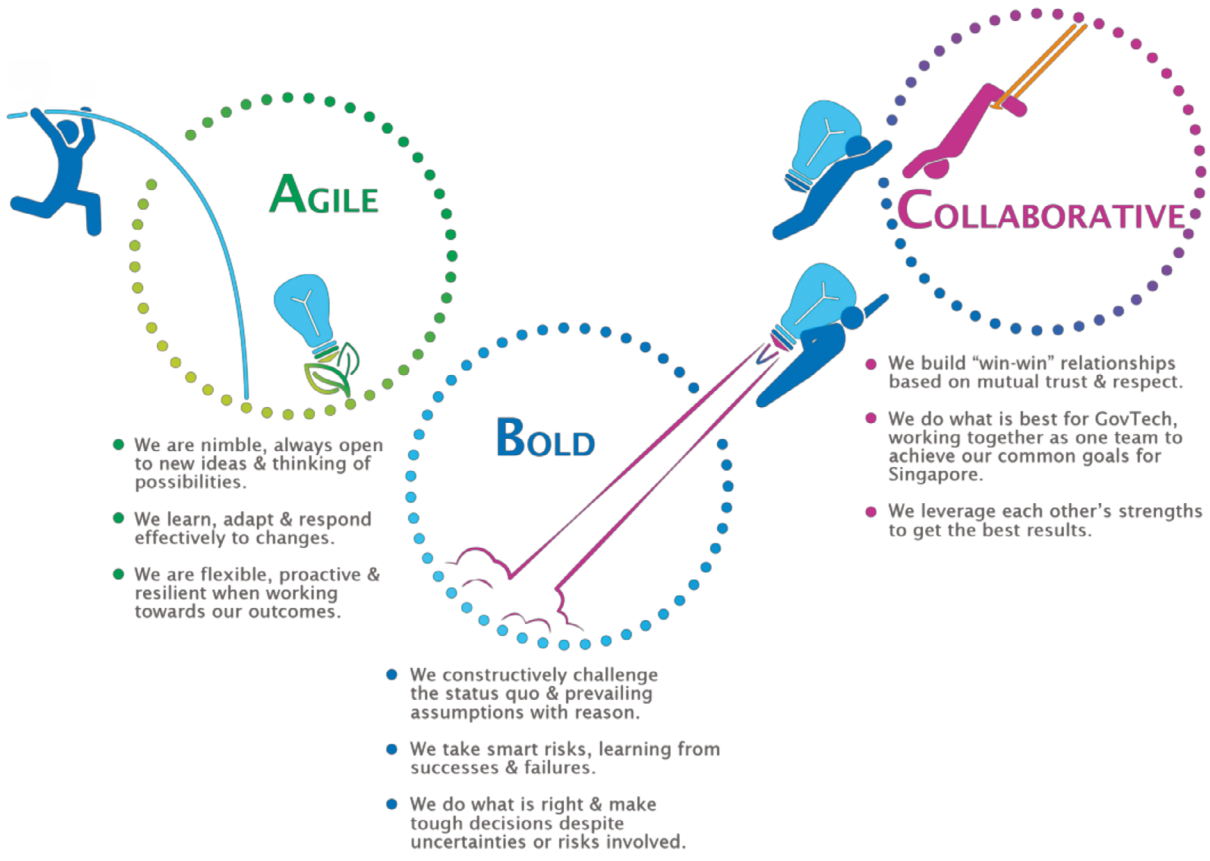
As the tech-related projects piled up, the team faced the challenge of securing the capability needed to meet demand. Some highly paid engineers did not see the attraction of working for the government, while others were already dispersed through different agencies. To support its recruitment drive for workers with the necessary skills, GovTech employed an unusual method: coding competitions. The top three winners of a HackerTrail challenge (a technology recruitment company) won the opportunity to apply their tech skills to contribute to the public good.

The Data Science Division is passionate about data-driven policymaking, as opposed to policy-driven *data-making*. The latter occurs when people generate data for the purpose of backing up policies they have already made. Accomplishing the former required a multidisciplinary effort, bringing together a diverse group of people who had been given permission to work in a different way. The team needed computer scientists who could code and build apps, social scientists able to ask the right public policy questions, and designers who could present

information in compelling ways—and the passion to dive deep into the data united them.

The combination of culture, flexible arrangements and benefits, good remuneration, and career development opportunities has proven successful in luring talent. GovTech's approach is evident from its three main traits, which are now stamped on its t-shirts: agile, bold, and collaborative. Being agile is about being able to pivot, to flexibly get things done. Agility also means moving fast without the formality usually found in government. Being bold means having the courage to experiment. And most importantly, no one person can make it happen. A collaborative culture imbues the organization, and many people join it for that reason.

Figure 4. Culture Values at GovTech



Source: GovTech

Evidence-based projects can be built in a variety of formats. If a government agency seeks insights into a policy question, data scientists can present their findings in a slide deck, for example. But a slide deck would not help in addressing an operations-related problem. In this case, data scientists might build dashboards or decision-making tools that can be used on a daily basis. If the issue concerns digital services for citizens, data-driven app design and customization of the service to the target audience becomes important. Ideas for digital services come from a range of sources, including formal or even informal coffee chats between staff from the agencies involved.

GovTech assists agencies with the development of new services or the enhancement of existing ones. It works to bridge the gap between policymakers and information technology departments, which often work in separate silos. The challenge usually lies in figuring out how to get people from diverse backgrounds to talk to one another and work as a cohesive team.

When approaching a data science project, GovTech teams soon learned that three important ingredients needed to be in place. The first was a clearly defined problem statement that articulated a project's potential impact at the outset. The teams put a lot of effort into working with agencies to scope data science projects. This involved: sharpening vague problem statements, such as "we wish to use AI to improve service delivery," to specific lines of analysis with well-defined metrics. The second ingredient was whether good-quality data could be made available to the GovTech teams. Without access to feedback and data in the right format, there is little for data scientists to work with. And having good-quality data somewhere in the government will be useless without the buy-in of stakeholders to hand it along the chain from the problem owner and data owner to the problem solver. It is never an easy task to get them to willingly engage with any frequency or openly share relevant information to move things forward.

With clarity in vision, access to good-quality data, and strong partnerships, the teams adopted a "fail fast, learn fast" mentality—the third ingredient. This was useful for iterating data science projects and deciding which ones would deliver the most value and should be prioritized. The biggest challenge was finding the ones that had the most potential to make a real-world impact, so the teams developed an evaluation framework to select data science projects with high impact potential. The framework entailed collecting use cases, building pilots around the cases within two months, and trying to demonstrate the potential value of the method. If the method proved useful, the teams would build a front-end interface to wrap around the core technologies and make it available to civil servants. If not, the teams would move on to something else. Although this framework cannot guarantee real-world impact from every data science project, the team's discipline helped them define which projects had the most potential to do so and ensure each project's relevance to the public sector.

Building data science workflows for public servants was just one part of the equation. Supporting the uptake of these solutions by civil servants turned out to be equally critical. The teams needed to repeatedly work with agencies to tailor analyses to produce actionable insights and create front-end data science products so that staff who do not know how to code were also able to use the tools in their work. The teams' collaboration with the Housing and Development Board is a good example of these practices. GovTech applied data science methodology to analyze 100,000 emails sent by the public to the board. The analysis revealed a topic on homeowners' questions about collecting their keys from the board, so the teams created an online portal and changed its collection process. Rather than assign collection dates, the board now allows homeowners to select a date to collect their keys. The data scientists then packaged their code into a front-end user interface so board staff could run the analysis on their own. By putting these data science tools into the hands of every worker, the government can better understand what citizens care about.

Key to GovTech's modus operandi is also its funding model, which borrows from the start-up world. Rather than going through a lengthy process of applying for project funding, each capability center within GovTech, including data science and AI, has its own pot of money for "white space" development. GovTech, rather than being a cost center for government, also generates its own

revenue. It made US$269.5 million in its first six months of operation, according to its first financial report.

Within GovTech, training in data analytics has become a priority too. To deepen its technical capabilities for information and communications technology (ICT) and Smart Systems, the agency created a Centre of Excellence that houses capability centers such as Data Science and AI, ICT Infrastructure, Application Development, Sensors and IoT, Cybersecurity, and Geospatial. In addition, the Smart Nation Fellowship Programme brings together talents from the tech industry to collaborate with public agencies and build meaningful digital solutions that improve the lives of citizens. The government has also committed to having a digitally competent workforce. By 2018, around 4,000 public service officers had training in data analytics, and the goal is to train 16,000 more.

What Was Achieved

GovTech works on the development of smart, data-driven applications to enhance or deliver more customized experiences for citizens. GovTech was the engine behind the creation of popular government apps such as:

- ***Beeline,*** is described as an "Uber for mini-buses." The Beeline team analyzed data to seed initial bus routes and continues to refine the routes through crowdsourced data.

- ***MyResponder,*** which was developed in collaboration with the Singapore Civil Defence Force, alerts volunteers to emergencies in their vicinity. For example, volunteers are able to assist in performing CPR or helping to apply a defibrillator are notified of cardiac arrest cases within 400 meters of their location.

- ***OneService,*** which was jointly developed with the Municipal Services Office, allows the public to share feedback on municipal issues.

- ***SingStat,*** provides a range of statistical data about Singapore.

GovTech's data scientists have also collaborated with SingHealth to use machine learning to analyze electronic medical records and identify patients at high risk of hospital readmission. The algorithm, which was published in an academic journal (Low, Lee, Ong et al., 2015) was about 80 percent accurate.

Another achievement of the data science team was customer segmentation for government use cases. For example, most librarians would know that their customer base includes young families, teenagers, and senior citizens. However, a deeper analysis of library data (e.g., age, loans, type of books borrowed) on borrowers revealed two distinct groups of senior citizens: the "retiree hobbyists" who regularly borrowed books on various leisure pursuits and grandparents looking for books with their grandchildren. This insight helps libraries better plan their collections and cater their activities to the different groups of senior citizens.

GovTech also revamped the data.gov.sg website to make data more accessible and help in communicating more effectively with data. The new site places heavy emphasis on data visualizations to allow users to understand the data at a glance and has a data blog for data narratives and stories. The data.gov.sg team will also be focusing on improving data quality and releasing more data streams, or application programming interfaces, to enable software application development on top of the data.

Case 3: Ecosystem Model, Mexico

Building a Data Analytics Ecosystem

Photograph from Government of Mexico

Highlights

Relative to the previous cases, Mexico is larger and less developed. Nevertheless, it is equally committed to its open data policy and it has created an ecosystem to foster data-driven policymaking. This multistakeholder approach relies on collaborative efforts and brings together industries, academia, and civil society organizations, in line with policy and political priorities. The first pilot data analytics initiative, Datalab, enabled a quick delivery of policy wins and raises awareness on the relevance of new technologies and the strategic use of data for government performance. What started with just three projects has flourished into countless initiatives and collaborations between government and different institutions.

Although sustained by an ecosystem of actors, the Mexican government was key to gaining the necessary environment of solid political leadership and support, capacity building strategies, and multidisciplinary teams. Another key issue in Mexico's success was its focus not only on economic gains but its emphasis on values connected to policy priorities. Data-driven solutions to public interest topics, such as anti-corruption and human rights, have gained widespread support among the main stakeholders.

Mexico's model is called an ecosystem model because the government's role was primarily one of coordination for a vibrant, data-literate civic society that was able to offset the government's limited resources for hiring or buying analytics delivery capability. The government engaged private sector partners who were keen to work with the government on projects and train public servants.

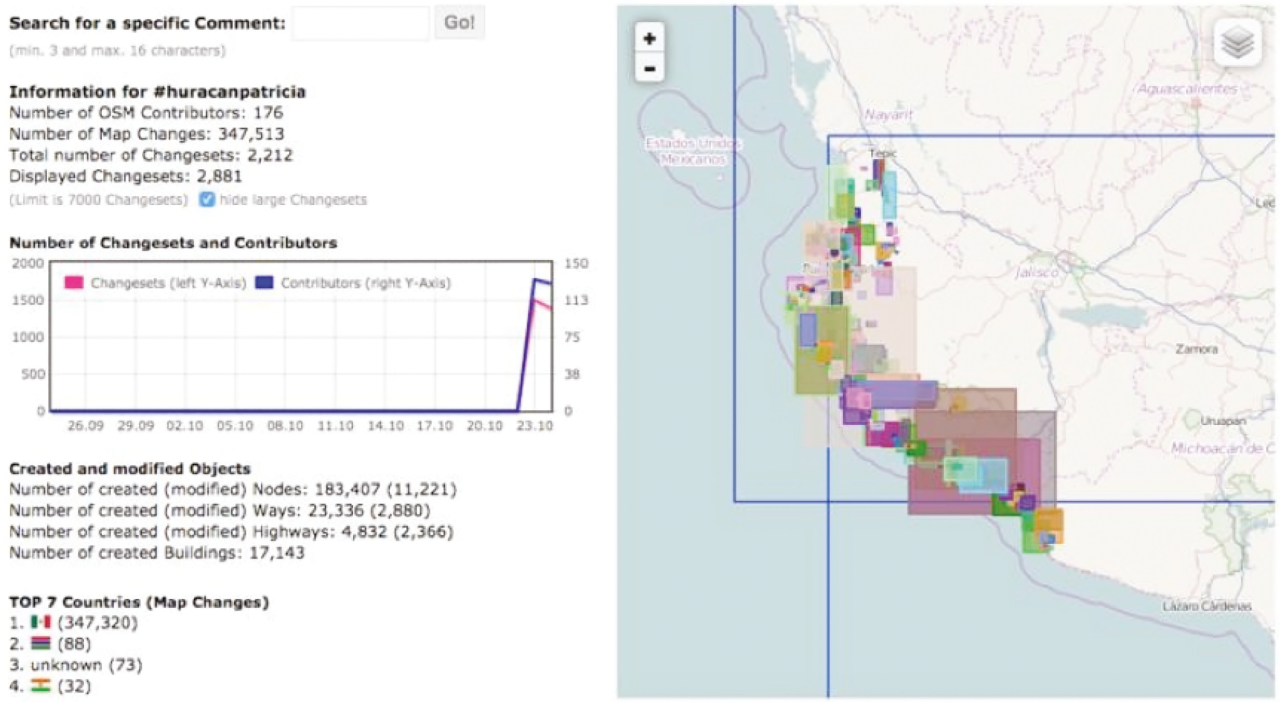
What Was Done

Since 2013, Mexico has been firmly committed to digital transforming its government, and it has the high-level political support to do so. Toward this end, Mexico has successfully established its central open data portal, datos.gob.mx, and developed technical and regulatory guidelines based on international open data principles. These measures have resulted in the gradual publication of data by all levels of government. Currently, 200 public institutions have published over 12,000 databases generating over 300,000 downloads.

This open data effort was tested to its limit in October 2015, when Hurricane Patricia—the strongest hurricane ever recorded in the Western Hemisphere—made landfall near Cuixmala. Using previously open and available data, critical information taken from a myriad of sources was linked together. The data included demographics, health facility locations, power infrastructure, road networks, airports, topography, and steady updates from international organizations regarding the hurricane's location, storm path, and speed. Armed with 38 sets of critical databases, some 400 global volunteers mapped 9,000 kilometers of roads and 90,000 buildings and hotels and potential mudslide zones, providing over 1 million people with vital information in real time.

As an example, the Humanitarian Team of OpenStreetMap (OSM) created a map to coordinate international efforts to capture and generate public information essential for addressing the emergency. In 72 hours, the team captured data on thousands of kilometers of roads and tens of thousands of construction areas in the coastal zone. Later, this task would be supported with the release of satellite images by DigitalGlobe and the creation of a data repository by the Office for the Coordination of Humanitarian Affairs of the United Nations in the Humanitarian Data Exchange (HDX).

Figure 5. OSM Changesets (last 30 days)



Source: OpenStreetMap

Hurricane Patricia provides a dramatic, real-life example of how open data can improve or even save lives on a potentially massive scale. However, the Mexican government's approach to opening public data—and setting up an ecosystem of different organizations to make the most effective use of that information—is also about bringing a broad range of innovative services to citizens. It was necessary to develop the capacity to collect, share, and analyze massive amounts of data in order to apply an exact methodology to the design of public policies. The government's human capital was capable of seeing social problems from the point of view of public-policy design but no training to carry out such an analysis from a computer's perspective.

Convinced of open data's potential, Mexico was ready to transform data into public value and impact. In June 2016, the government launched the Datalab, an initiative implemented by the General Direction of Open Data with the main goal of promoting evidence-informed policymaking driven by data science and enabled by digital technologies.

The Datalab was a collaborative effort between the Coordination of National Digital Strategy and other federal agencies as well as academia (namely the National Public Policy Lab at the Center for Economics Research and Teaching [CIDE]), students, and graduates with a background in data science, applied mathematics, or other related areas. The initiative centered on the conception of public sector institutions as data users, aiming to reduce the gap in data expertise within the public sector to foster data-driven policymaking.

When created, the Datalab prioritized the administration of a survey across selected public sector institutions, including the ministries of health, public administration, and social development in addition to the National Institute of Ecology and Climate Change, and the Mexican Institute of Youth. The survey assessed the state and capabilities of these institutions regarding open data publication and their potential use of data given their policy priorities in order to identify how digital technologies could contribute to achieving those priorities.

The Datalab initially worked on three main projects:

- Representatives from the Ministry of Health worked with the fellows and professors at CIDE to develop a dashboard that provides timely information on health performance indicators, which would show the value of real-time data for public sector productivity. The cohort also designed advanced data-driven projections and simulations to inform public officials on how timely policy interventions could reduce teenage pregnancy rates in the midterm and long term.

- The Ministry of Public Administration worked with the Datalab to explore how data analysis could support the development of an automated system for the prevention, detection, and reduction of conflicts of interest and of nepotism in public procurement and civil service appointments.

- he third cohort, which involved the Ministry of Interior, focused on developing a machine-learning algorithm to analyze media coverage on aggressions against journalists and human rights defenders in Mexico. The objective was to build a database from unstructured information using natural language analytics to identify risk factors affecting the work of media rights defenders. The project seeks to understand the social dynamics and violence in the country, identifying the factors that endanger human rights defenders and journalists in the exercise of their activities. In this way, the cohort can identify the places of greatest danger as well as test hypotheses about the causes that trigger violence against these actors. The biggest challenge was the great diversity of information and its sources, mostly from journalistic notes. The team had to build a program that learned to read the news, identify reports that are about assaults against journalists and activists, and extract essential information from the article.

The Datalab enabled the government to deliver quick policy wins and raise awareness about the relevance of new technologies and the strategic use of data for government performance. But as a one-time initiative, the Datalab did not ensure the continuity of capacity-building efforts across the public sector. To ensure continuity and sustainability, the Coordination of National Digital Strategy created the Goblab in May 2017 as a permanent initiative within the Office of the President, building on the original team that formed the Datalab.

The data chapter of Goblab aims to design data-driven, digital, innovative solutions to the Office of the President and other public sector agencies covering policy areas such as sustainable development, human rights, climate change, anti-corruption, and the security agenda. The media chapter, on the other hand, experiments with policy perception, citizen participation, and social media in order to provide a stronger evidence base for decision-making and policy priorities included on the president's agenda.

By the end of 2017, Goblab was piloting projects in collaboration with the Ministry of Public Administration for the new anti-corruption system, a new food security index to foresee and prevent food shortages in collaboration with the Ministry of Social Development, a new Human Rights Indicators System with the Ministry of Interior, and the use of open data to follow up on prosecutions cases and missing persons. These projects were developed with the participation of the United States Agency for International Development, the Inter-American Development Bank, and academic institutions such as El Colegio de Mexico and Mexico's National Autonomous University.

How It Was Done

The achievements of Mexico's open government data policy to date result from a favorable context comprising different policy elements, including:

High-level political commitment and solid political leadership. The Goblab was the first formal data and media laboratory for policymaking within the president's office. As part of the National Digital Strategy, the team had complete support from the highest level of government.

Capacity-building support and technical guidance. These elements favor increasing institutional buy-in, awareness, and skills development across the public sector. With limited resources available, the Datalab had initially hired just six data scientists (as a result of an open call that brought in more than 120 applications), each with a civic background to work as fellows for four months. The focus then shifted to training. More than 3,000 civil servants were trained in data analysis in the past year, as the Mexican government is pursuing this goal aggressively. Beyond the efforts and initiatives implemented at the center of government, there were more initiatives aiming to build data skills among public sector institutions led by other actors, namely:

- The Research Center on Geospatial Information Science (Centro GEO) launched an online course to build capacities on the reuse of geospatial data. The course was available through MexicoX, the eLearning platform of the Ministry of Public Education that was launched in 2015 in collaboration with Coordination of National Digital Strategy.

- The Ministry of Finance, in collaboration with the Office of the President, launched a massive online open course on the use of open budget data published by the ministry. The course targeted mainly public officials, as it was linked to the development of the evaluation and performance indicators of the finance ministry. Roughly 17,500 people signed up for the course, and 3,900 people completed it.

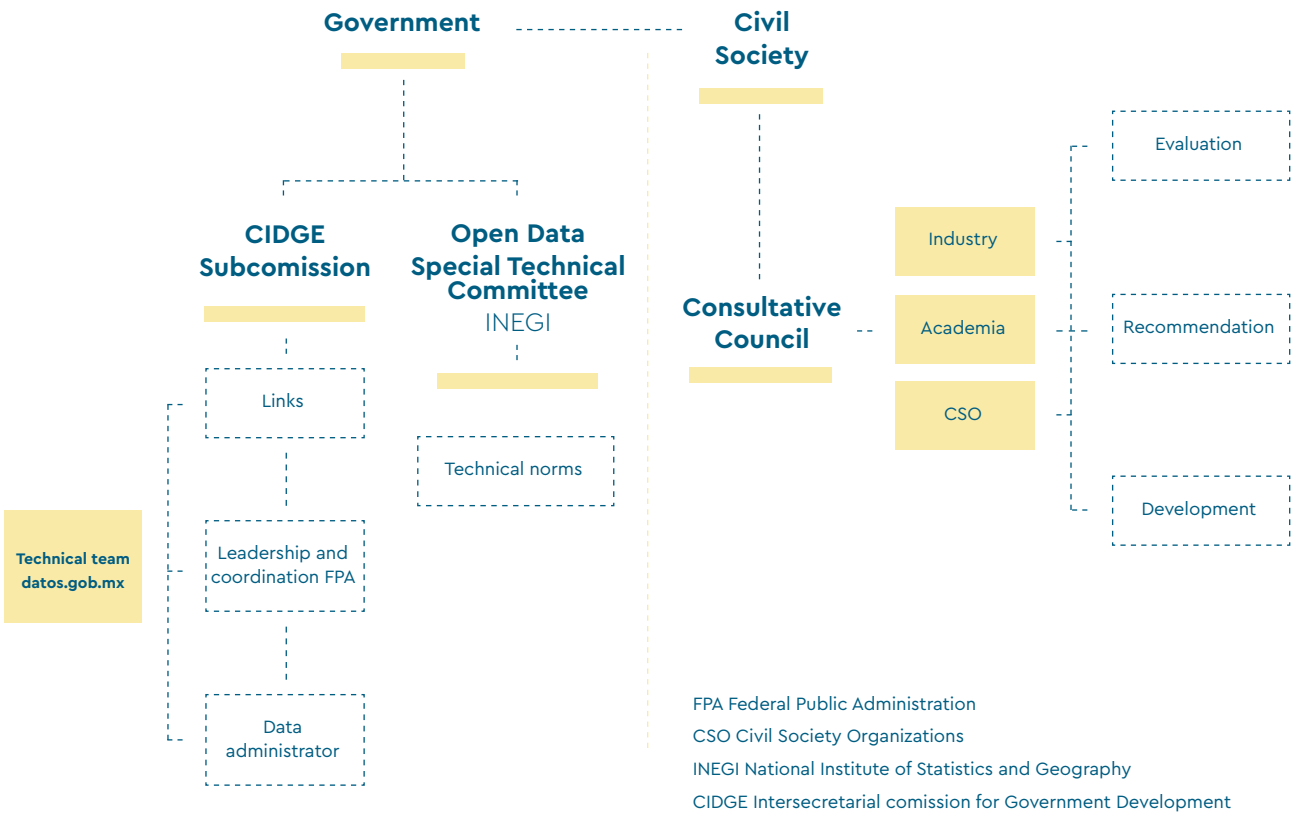
- The General Direction of Open Data is working on the development of an Open Data Competency Standard with the help and guidance of CONOCER, the National Council for Standardization and Certification of Labor Competencies. CONOCER, part of the Ministry of Public Education, is responsible for providing competencies certification, training, and evaluation in line with the priorities from the private, civil society, and public sectors.

- In 2018, as part of the new Open Data Policy Implementation Guide, the General Direction of Open Data undertook monthly capacity-building webinars for public institutions.

A multidisciplinary team. The members included data scientists, designers, communicators, and human rights experts. A flexible and dynamic human resource management at the Office of the President allowed new hires with new skills who favored data-driven innovation. Within Goblab, potential solutions for policy issues are engineered based on the mix of digital skills and experimental approaches (e.g. design thinking, user-centric design, machine learning).

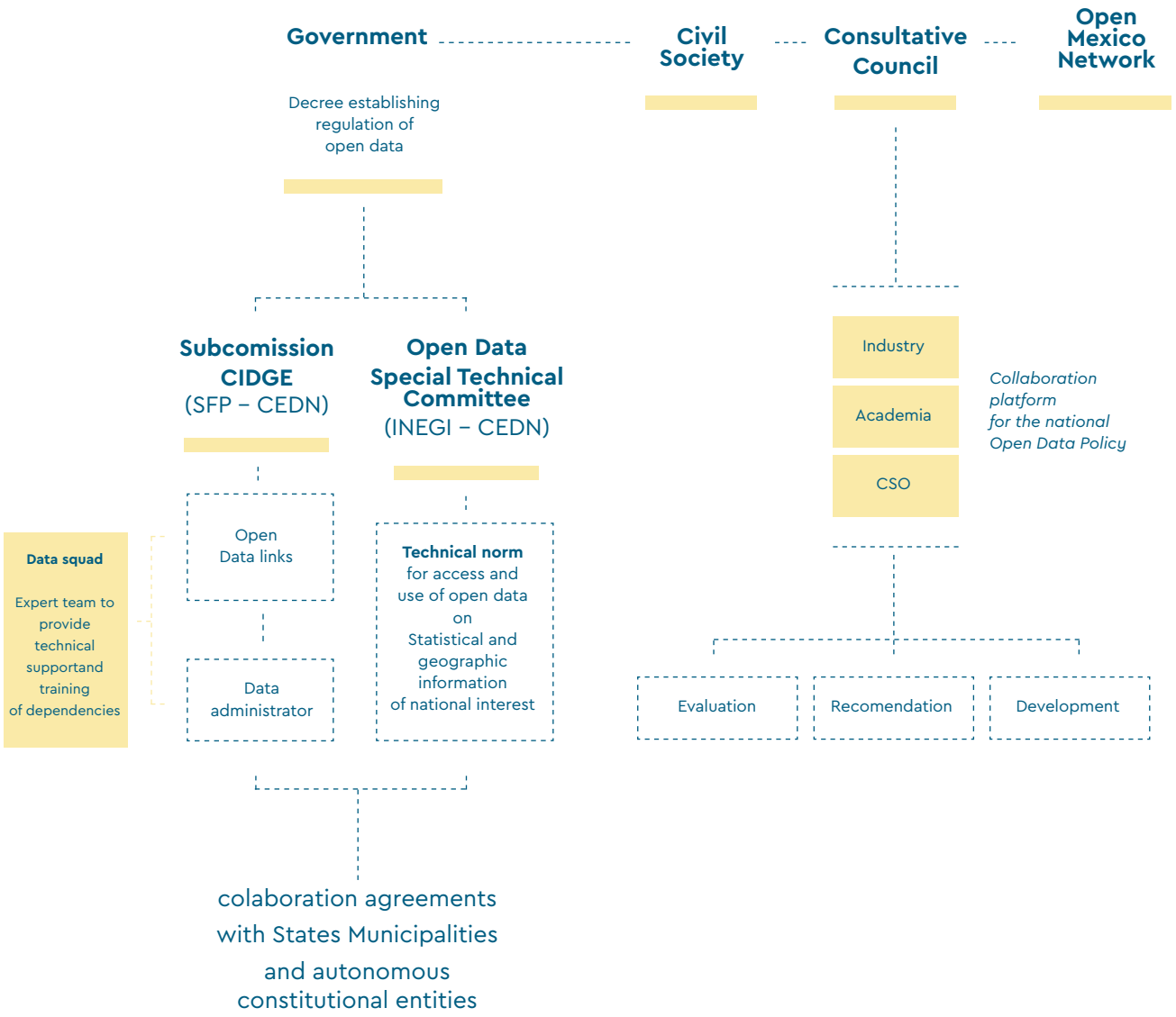
A multistakeholder approach. Goblab projects show the willingness of the Mexican government to foster data-driven public sector innovation with a multistakeholder approach that relies on collaborative efforts and brings together a number of public entities, in line with policy and political priorities. Projects are carried out collaboratively with international organizations, civil society organizations, academics, and the telecommunications and IT industries. Each of the initiatives can be enriched in work groups, committees, and evaluation sessions with representatives from each of these sectors. In the same way, the establishment of alliances allows for the discovery of areas of opportunity to improve and realize digitization plans.

Figure 6. Mexican Data Governance



Source: México Digital (2018)

Figure 7. Institutional Structure



Source: México Digital (2018)

Central policy and clear policy goals. The National Digital Strategy is the government's action plan to build a Digital Mexico. Its fifth objective is "Civic Innovation and Citizen Participation," which is designed to:

- Promote civic innovation to solve problems of public interest through ICT
- Use data for the development and improvement of public policies

- Generate citizen complaint tools and applications on multiple platforms
- Prevent and mitigate the damage caused by natural disasters through the use of ICT

The initiatives focus on economic gains as well as social and good governance values connected to policy priorities in the country (e.g., anti-corruption and human rights). This focus was driven by the team's culture rather than its design. Also, Digital Inclusion, a cross-cutting initiative, ensures that digital services and applications reach the vulnerable sectors of the population and have a real impact on their lives. It focuses on health and financial inclusion in particular.

What Was Achieved

Although the Goblab team is relatively new, the Mexican experience in data analytics has already achieved visible results. Many of these are more intermediate than more established data analytics teams' (such as New York City's). Among them, the following stand out:

Red México Conectado, is a multistakeholder mechanism to promote best practices, build capacities, and enable the publication of open data from the local level in datos.gob.mx.

In collaboration with Goblab, the government of Mexico developed the project Open Data 100 MX, the first-ever mapping of companies using open data as a key input to generate economic and social value. More than 100 companies were identified, many of which are new businesses entirely based on open data or using open data as one of their core business proposals.

As part of the National Digital Strategy, the Goblab team has partnered with the Open Data Institute, DEMOS (a joint effort of PIDES Social Innovation –today C Minds – and Dev.F), and the United Kingdom's embassy in Mexico on Labora, which is a platform that supports civic and social entrepreneurs by offering capacity-building trainings and mechanisms to connect them to a world-class network of businesses, mentors, and investors. Labora is designed to accelerate the entrepreneurs' impact through the use of open data in their products and services. The project is already in its second phase, where, in addition to continuing to build capacity to develop their businesses into small and medium enterprises, the entrepreneurs will work with sector incubators and accelerators—starting with the fintech sector—to make open data a productive input for the economy.

Together with University of Chicago, Goblab is working to reduce maternal mortality in Mexico. Though the project is in a preliminary stage, the combined team is studying the available data at a more granular level than it has been studied before to better understand why the maternal mortality ratio has remained relatively stagnant over the last 10 years. Despite poor-quality data and disjointed and unlinkable datasets, the data are still useful to data science or even machine-learning methods. While unable to develop an individual longitudinal dataset to apply classification approaches on, the team seeks to apply aggregate regional, local, and municipal models. In looking at factors such as socioeconomics, available facilities, access to facilities, urbanicity, and location, the team will apply various models to identify the highest-risk regions and the key indicators of risk. It will soon be leveraging decision trees, random forests, logistic regression models, Poisson regression models, and hierarchical models to various capacities in this regard.

The 2017 Open-Useful-Reusable Government data Index ranks Mexico in second place for reuse and creation of impact with data. This Organisation for Economic Co-operation and Development index also puts Mexico at fifth place in the global table, five places higher than in 2015.

Mini Case:

Multilateral Partnerships in Indonesia

Pulse Lab en Jakarta (PLJ) is a joint initiative of the United Nations and the Government of Indonesia, via United Nations Global Pulse and BAPPENAS, the Ministry of National Development Planning. The PLJ combines data science (big data, real-time analytics, and artificial intelligence) and social research to accelerate the discovery and adoption of data innovation for sustainable development and humanitarian action.

The Lab provides three main services:

- Conduct cutting-edge research on a blend of new and existing data sources
- Generate data products and tools to identify solutions
- Connect data innovators from across the region and scale collective impact

Pulse Lab relies on a core team of 28 people, including experts in data science, data engineering, statistics, policy, and ethnography. But many of the lab's research projects are delivered in conjunction with its partners, including players from the private and public sectors, such as the government of Indonesia, United Nations agencies, foreign governments, data philanthropy partners, and civil society organizations. Partnerships may be established by:

- Becoming a data philanthropy partner to share your data for analysis
- Becoming a research partner to work on a project with PLJ
- Becoming a technology partner and building prototypes and/or test new tools
- Presenting a big data analysis tool or method at a training session or workshop
- Sponsoring a fellow to join the Pulse Lab to work on data visualization, data analysis, GIS mapping, and/or software development
- Providing funding or in-kind support for PLJ's work

Using datasets drawn from mobile communications, remote sensing, and social media, among others, PLJ has generated insights for policy and practice on topics ranging from fuel subsidies to natural disasters.

These are some of the PLJ's main projects:.

Improved public transport

PLJ collaborated with Jakarta Smart City on a project to enhance transport planning and operational decision-making through real-time data analytics. Using data from TransJakarta (the city's bus rapid transit system) buses and passenger stations, the project mapped origin-destination trends and identified bottleneck locations, information which can be used to identify whether new routes are needed. The project also explored the possibility of using real-time data to determine passenger waiting times in order to enhance the efficiency of the bus dispatching system.

Citizen feedback for local government

Like many middle-income countries, decentralization is a national priority in Indonesia. As their administrative responsibilities and fiscal resources have increased, local governments have sought better evidence to inform policy. In this project, PLJ explored the contribution of advanced data analytics to local government decision-making by generating insights from a combination of existing complaint systems and passive feedback from citizens on social media. The results demonstrate the potential utility of (a) near real-time information on public policy issues and their corresponding locations within defined constituencies, (b) enhanced data analysis for prioritization and rapid response, and (c) deriving insights on different aspects of citizen feedback. The publication of citizen feedback on public-facing dashboards can enhance transparency and help constituents understand how their feedback is processed.

Coming next: Disaster Mon

The PLJ team is currently building a cyclone monitoring prototype. The team is developing an integrated big-data analytics prototype to provide timely insights for natural disaster monitoring, emergency response, and management of cyclones, earthquakes, hurricanes, and floods in Indonesia and countries in

the Pacific region. Nicknamed "Disaster Mon," it relies on multiple data sets that are available globally through public application programming interfaces, in conjunction with other data sets that PLJ has access to through data partnerships. The current prototype features three data insight layers related to logistics planning, social media communication, and socioeconomic variables, which are based on multiple data sources.

Conclusions

This report has set out examples of three governance models used by data analytics teams. In reality, the distinctions between them are not often clear cut. We would expect future data analytics units to draw on lessons and elements from all of the models.

This is important to emphasize because setting up a successful data analytics team will not be as easy as a government simply choosing the model it prefers and sticking to that path. The early leaders in this field have recognized that the reality is inevitably a bit messier, and they have shown flexibility by adapting their ways of working according to their successes and failures as well as changes in political circumstances.

For governments that have a culture of centralization, strong political leadership behind a data-led and modernization agenda, and mechanisms or a mandate for driving change at scale across institutions, a centralized model akin to that used by New York City is likely to be the best place to start.

For governments that have a culture of collaboration, relatively deep and broadly distributed data capability, and a preference for using distributed funding models that give departments responsibility for prioritizing spending, adopting the federated model Singapore began with is an appropriate starting point.

For governments that have a vibrant, data-literate civic society, limited scope to hire or contract analytics capability for the public service, and a private sector keen to partner with government on projects, the ecosystem model used by Mexico would be the best place to start.

However, some degree of centralization and government responsibility for data analytics is inherent to all these organization models. If governments want to establish strong data analysis capabilities, their minimum investment must be in a function that is capable of coordination (and potentially also some delivery) of data analytics projects across state institutions. This is true whatever the government's size, scope, or level of capability. All the case studies cited in this report provide examples of where countries have been intentional, flexible, and committed to setting up their teams. Investment and political support have reaped rewards, and without providing both, governments risk failing to make the most of their data's potential to improve the quality of their policy and the lives of their citizens.

Bibliography

Basu, M. 2019. "West Kalimantan launches open govt vision." GovInsider, March 7. https://govinsider.asia/data/west-kalimantan-launches-open-govt-vision/?ct=t(GI_PVSub381.1_COPY_01)&mc_cid=f362963742&mc_eid=743ca905ba.

Chambers, J., y C.J. Lin. 2018. "Exclusive: Inside Indonesia's Plans to Fix Corruption and Inequality." *GovInsider*, November 16. https://govinsider.asia/smart-gov/inside-indonesias-plans-to-fix-corruption-and-inequality/.

"Data Revolution Advisory Group Named by UN Secretary-General." 2014. UN News, August 29. https://news.un.org/en/story/2014/08/476242-data-revolution-advisory-group-named-un-secretary-general

Glover, J. 2018. "How to Do Data Analytics in Government." Government Technology, July 16. http://www.govtech.com/data/How-to-Do-Data-Analytics-in-Government.html

GovTech. 2019. "GovTech Organisation Chart." Our Team, GovTech. Updated May 24, 2019. https://www.tech.gov.sg/who-we-are/our-team/.

Lane, A. 2017. "How Governments around the World Are Turning to Data." *Innovation Enterprise*, May 10. https://channels.theinnovationenterprise.com/articles/how-governments-around-the-world-are-turning-to-data

Low, L.L. Low, K.H. Lee, M.E.H. Ong et al. 2015. "Predicting 30-Day Readmissions: Performance of the LACE Index Compared with a Regression Model among General Medicine Patients in Singapore," *BioMed Research International* 2015.

Mexico Digital. 2018. "¿Cómo Trabajamos?" *Mexico Digital (blog), Government of Mexico*. November 28, 2018..

Muente Kunigami, A., and M. Marshall. 2018. "Es la hora de datos más abiertos en América Latina." *Abierto al público (blog), Inter-American Development Bank*. August 9, 2018. https://blogs.iadb.org/abierto-al-publico/2018/08/09/datos-mas-abiertos-en-america-latina/.

NYC Analytics. n.d. "About Open Source Analytics." *Moda Project Library*. https://moda-nyc.github.io/Project-Library/analytics/. ———. 2013. "NYC by the Numbers" in Annual Report 2013. New York: NYC Analytics.

OECD (Organisation for Economic Co-operation and Development). 2018. Open Government Data in Mexico: The Way Forward. OECD Digital Government Studies. Paris: OECD Publishing. https://doi.org/10.1787/9789264297944-en. Open Data Barometer. n.d. "The Open Data Barometer." Open Data Barometer. https://opendatabarometer.org/?_year=2017&indicator=ODB

Prime Minister's Office, Singapore. 2017. "Formation of The Smart Nation and Digital Government Group in the Prime Minister's Office." Prime Minister's Office, Singapore. Updated December 24, 2018. https://www.pmo.gov.sg/newsroom/formation-smart-nation-and-digital-government-group-prime-minister's-office

Santiso, C. 2018. "The Open Data Revolution Is Sweeping the World. Has It Reached the Americas?" *Apoltical*, September 18. https://apolitical.co/solution_article/the-open-data-revolution-is-sweeping-the-world-has-it-reached-the-americas/.

Siddons, E. 2018. "The Global South Is Getting Its Own AI Healthcare Revolution." Apoltical, September 18. https://apolitical.co/solution_article/the-global-south-is-getting-its-own-ai-healthcare-revolution/.

Young, A. 2018. "The GovLab Index: Open Data – 2018 Edition." *GovLab (blog), GovLab*. September 26. http://thegovlab.org/open-data-index-2018-edition/


IDB
Inter-American
Development Bank