

Replicating Urban Analytics Use Cases: Conceptual Shortcomings, Implementation Challenges, and Recommendations for the Field

Craig Campbell
Civic Analytics Network

January 2019



Replicating Urban Analytics Use Cases: Conceptual Shortcomings, Implementation Challenges, and Recommendations for the Field

Craig Campbell
Civic Analytics Network

January 2019



ABOUT THE AUTHOR

Craig Campbell is the Assistant Director for Policy & Operations for the NYC Mayor’s Office of Data Analytics (MODA). Prior to working for the City of New York, Craig researched trends in municipal data analytics, supporting several national policy networks and research programs at the Ash Center for Democratic Governance and Innovation at Harvard Kennedy School. Craig holds a degree in architecture and mathematics from Amherst College.

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.

CONTENTS

Acknowledgments	iv
Introduction	2
Analytics Replication as Urban Policy Diffusion	3
Data Analytics' Replicability Problem	4
Adapting an Analytics Use Case	5
Evaluating Consistency of Implementation Environments	6
Use Case Consistency Template	7
Recommendations for the Field	9
Recommendations for Urban Analytics Practitioners	9
Recommendations for Networks of Analytics Practitioners	10
Recommendations for Future Research	11
Conclusion	12
Case Studies	14
New Orleans Smoke Alarm Distribution	14
Predictive Policing Software	16
Appendix A. Inspiration and Implementation Vehicles	19
Inspiration Method: Peer Learning Communities	19
Inspiration Method: Clearinghouses	20
Implementation Vehicle: Pro Bono Technical Assistance	21
Implementation Vehicle: Procurement	22

ACKNOWLEDGMENTS

This project would not have been possible without the leaders in urban analytics who graciously agreed to be a part of this project: thanks to Sam Edelstein, Danielle Fulmer, Brian MacDonald, Sascha Haselmayer, Andrew Nicklin, Carter Hewgley, Laura Meixell, Robert W.S. Ruhlandt, Gaylen Moore, and Tamas Erkelens for their comments and review. The author is especially grateful for the support of Data-Smart City Solutions Project Manager Katherine Hillenbrand, the thought leadership of former San Francisco Chief Data Officer Joy Bonaguro, and the guidance of Harvard Kennedy School’s Daniel Paul Professor of Practice, Stephen Goldsmith.

The Civic Analytics Network and its outputs, including this paper, are generously funded by the Laura and John Arnold Foundation. This paper is an independent work; the views expressed herein are those of the author and do not necessarily represent those of the funder.

The 2010s have seen a rapid expansion of executive-level data and analytics positions in local government. Following the rise of the “chief data officer” in both the private and public sectors, and what many see as inadequate urban policy leadership at the federal level, expectations run high for the impact data analytics could make in cities. To help fulfill this promise, a patchwork of universities, philanthropies, and non-profits are supporting cities in adapting analytics innovations developed in one city to others across the country.

How does an algorithm that prioritizes smoke alarm distribution in New Orleans change the way a similar service is delivered in Syracuse? Why does a statistical model for rodent mitigation developed in Chicago fail to work in Pittsburgh? How should a data-processing solution in Louisville, or a problem-solving methodology in New York City, be iterated in other cities? This paper—developed from academic literature review, interviews with analytics practitioners in local governments, and notes from roundtables of experts—explores both the theoretical solvency and practical considerations for replicating data-analytics use cases from city to city.

The findings are mixed. Stories of analytics replication are often told in terms of their potential rather than realized value. Both “analytics” and “replication” suffer from loose definitions, and applications of data science in cities—even in the largest cities with the most sophisticated analytics programs—are often proofs of concept, rather than a core component of performance management and innovation delivery. What is clear is that data analytics solutions are far from a silver bullet to public problems, and there is substantial work that must be done to develop the way analytics “successes” are quantified before the field can have a productive conversation on inter-city analytics replication at scale.

The first section of this paper contextualizes the topic of civic analytics replication in the fields of technology transfer and urban policy diffusion, arguing that “adaptability” is a more appropriate conceptual framework for reconfiguring analytics use cases in new contexts. A template of factors to consider when attempting to adapt a project from one city to another follows. Brief case studies illustrate the potential and challenge of adapting analytics projects, and the paper concludes with suggestions on types of projects to adapt and recommendations for the field.

INTRODUCTION

In 1975, urban sociologist Janet Rothenberg Pack published a major study of the adoption of an advanced statistical model in regional planning agencies across the United States. Pack distributed surveys to 1,500 regional planning agencies and conducted site visits at 18 of them. Her results indicated that adoption of the model—which was designed to assist in decisions about where to place sewers, invest in transportation infrastructure, and constrain land use, for example—was common and that the model had been configured successfully in many unique jurisdictions.¹

“But what of actual model use?” Pack asked five years later, when she returned to those same agencies to assess whether their adoption was worth the effort. Were the models “incorporated into the routine analytic apparatus of the agencies? For what purposes are they used?”²

Where the models were still active, she found, they were being used to answer only narrow questions or to produce summary statistics for federal reporting requirements. More in-depth policy analysis and simulation of various planning scenarios—the hallmark uses of the tool that spurred its adoption in the first place—were “almost entirely absent.”³ A powerful data-analysis tool, widely recognized as applicable across regions, had been created and dispersed, but “replication” had failed at the level of implementation and long-term maintenance.

At a 2016 meeting of leading municipal analytics practitioners and experts at the Harvard Kennedy School, Johns Hopkins GovEx’s then-director of advanced analytics, Carter Hewgley, assessed the opportunities for analytics replication: “The good news is that problems and opportunities in U.S. cities are similar, meaning there is unending replication potential,” he said. The bad news was that lack of good protocols for use case discovery, challenges accessing and standardizing data, and uneven investment in data-literate human capital make analytics use cases difficult to generalize and

.....
1 Janet Rothenberg Pack, *Urban Models: Diffusion and Policy Application*, Issue 7 of Monograph Series (Philadelphia: Regional Science Research Institute, 1978).

2 Janet Rothenberg Pack, “The Failure of Model Use for Policy Analysis in Regional Planning,” *Systems Analysis in Urban Policy-Making and Planning*, ed. Michael Batty & Bruce Hutchinson (New York: Plenum Press, 1983): 293.

3 Ibid.

import into different cities. At a time when the value of predictive analytics is widely recognized as a tool for better decision making and “chief data officer” is an increasingly common title in municipal government, cities still face the same challenges adopting analytical models into routine operations they have faced for decades.

Analytics Replication as Urban Policy Diffusion

In the early 20th century, Supreme Court Justice Louis Brandeis famously validated local policy experimentation with his characterization of *federalism*, a structure for public administration in which individual regions govern themselves independently within a larger federal union. This system encourages policy innovation, Brandeis writes, by allowing “a single courageous State [to], if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country.”⁴

The same logic Brandeis theorized for American states applies to cities. In the past century, municipalities large and small have operated as “laboratories” of such policy innovations as participatory budgeting, inclusionary zoning, living-wage laws, and antismoking measures.⁵ Cities learn from, or at least influence, one another without intervention by a larger state or federal body.

The Information Age flattens the learning curve. Computer code is easily sharable, making data analytics (broadly defined, the use of mathematical models and data-processing technology to inform decisions and optimize business processes) an especially compelling subject for inter-city learning.

At a closer look, however, “analytics” proves to be a slippery object of inquiry. An analytics “use case” is more than an algorithm; it is a constellation of mathematical methods, technology capital, personnel skills, process design, and policy implementation.⁶ Transferring one is not so much a simple matter of sharing code as it is a

.....
⁴ New State Ice Co. v. Liebmann, 285 U.S. 262 (1932).

⁵ Charles Shipan & Craig Volden, “Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S. Cities to States,” *American Journal of Political Science*, Vol. 50, No. 4 (October 2006): 825–826.

⁶ Barry Bozeman proposes the “contingent effectiveness model” to describe the interdependency of artefactual knowledge, organizational capabilities, and situational factors in successful technology transfer across organizations. The model maps the relationships (contingencies) between various actors and instruments in the transfer

process of adapting a sociotechnical system to a new context. In interviews for this paper conducted with city employees who implement analytics use cases, all underscored that effectively replicating depends as much on a culture for evidence-based decision-making and the political will for change as it does on a model's degree of statistical significance.

Data Analytics' Replicability Problem

In an iconoclastic 2008 op-ed, *Wired* magazine's then editor-in-chief Chris Andersen declared the scientific method "obsolete."⁷ The increase in available data and computational power to analyze it, he argued, renders the rigor of the experimental process a relic of the past: "We can stop looking for models. We can analyze the data without hypotheses about what it might show."⁸

Andersen's comments reflect a foundational tenet of applied analytics: the question of *is it true?* is less operative than *does it work?* Data analytics, by nature, is an observational rather than experimental way of studying phenomena and behavior. As a field of practice, it owes more to hacker culture and management innovation than slow-moving randomized-controlled trials and program-evaluation procedures.

The very identity of "analytics"—a business term of art—is at odds with formal standards of academic rigor. This presents a methodological challenge for "replicating" analytics projects. Replicability is scientifically precise: the gold standard of accurate research is whether the findings of an experiment can be independently reproduced, should the experiment be performed again under the same conditions.

.....
process, including the *transfer agent*, the institution or organization seeking to transfer technology; the *transfer medium*, the vehicle, formal or informal, by which the technology is transferred; the *transfer object*, the content and form of what is transferred; the *transfer recipient*, the organization or institution receiving the transfer object; and the *demand environment*, meaning the market and non-market forces pertaining to the need for the transferred object. Barry Bozeman, "Technology transfer and public policy: a review of research and theory," *Research Policy* Vol. 29 Iss. 4-5 (2000): 626.

7 Further: "The scientific method is built around testable hypotheses. These models, for the most part, are systems visualized in the minds of scientists. The models are then tested, and experiments confirm or falsify theoretical models of how the world works. This is the way science has worked for hundreds of years . . . Petabytes [of data] allow us to say: 'Correlation is enough.'" Chris Andersen, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 23, 2008, <https://www.wired.com/2008/06/pb-theory/>.

8 Ibid.

Replication, as a conceptual framework, treats cities as generic experimental environments. It risks indulging a “technocratic replication fantasy” where models are viewed not as adaptable tools but “traveling solutions” in search of a problem.⁹ While municipalities have much in common, each has a unique physical, demographic, and administrative makeup. “Service delivery is anything but codified and varies wildly in local government,” commented former San Francisco Chief Data Officer Joy Bonaguro. “Rotely repurposing an algorithm is a flawed theory of the world.”¹⁰

To describe the process of reconfiguring a whole use case, not just reproducing an algorithm, using the vocabulary of *adaptation* and valorizing local conditions and expertise is more apt.

ADAPTING AN ANALYTICS USE CASE

Adapting a use case starts with *inspiration*, or the introduction of a novel idea into a new context. This requires the project in the originating city to be discoverable, whether through formal mechanisms, such as peer learning networks or a policy clearinghouse, or informal contexts, such as an open source code on a GitHub repository accessible through a web search.

Once local policymakers and program managers determine that the project is worth adapting, an *implementation agent* must manage the logistics of transferring knowledge capital such as data documentation, source code, case studies, and the know-how of practitioners in the originating city (through, for instance, a personnel exchange or tool demo). The implementation agent may be a pro bono partner such as a university, a company selling the analytical tool, or personnel in the city in which the project is being adapted.

For a more detailed overview of the institutions that consult on the replication of data analytics projects and the ways municipal analytics leaders discover their peers’ projects, see *Appendix A: Inspiration and Implementation Vehicles*.

.....
⁹ Jamie Peck, “Global Policy Models, Globalizing Poverty Management: International Convergence or Fast-Policy Integration?” *Geography Compass* Vol. 5 Iss. 2 (2011): 176.

¹⁰ Joy Bonaguro, phone interview with author, February 23, 2018.

Evaluating Consistency of Implementation Environments

The implementation agent, in collaboration with both the cities *from* which and *to* which the use case is being adapted, should develop an implementation plan that accounts for differences between municipalities' technical, political, and organizational environments.

Identifying what is the same and what varies in the problem definition, available data, and service delivery mechanism in each city will help indicate whether a project in one context is generalizable and help determine for how to reconfigure algorithms and other analysis methodologies to a new urban context. Each inconsistency identified should be accompanied by a specific tactic accounting for the variation. Some differences, such as inconsistencies in data structures, can be accommodated at a technical level when acknowledged upfront. Others, such as a lack of training and maintenance resources, may indicate replication is not the right solution to the local problem.¹¹

Though it lacks the scientific rigor of experimental replication, this process encourages intention and care for local context throughout the adaptation lifecycle. The template below can serve as a six-part checklist for evaluating the consistency of both the analytics product itself (e.g., statistical model, data) and the factors shaping how it can be used (e.g., cultural norms, service delivery mechanisms). Each step is accompanied by a set of key questions and illustrations of adaptation challenges for the recipient city and implementation agent to consider.

These steps roughly follow the steps of adapting a use case chronologically, but a full assessment of these factors should be conducted at the outset. This will focus project planning on the ultimate goal of improving service delivery, rather than simply implementing an algorithm for its own sake.

.....
¹¹ "Or it's just not the right time. So many projects just haven't ripened yet, but they will." Carter Hewgley, correspondence with the author, October 22, 2018.

Use Case Consistency Template

1. PROBLEM: Challenge or issue warranting public resources

- Is the issue a well-defined problem in the recipient city?
- Is the issue a priority among political executives?
- Are relevant departmental leaders and program managers interested in changing the status quo?

One chief data officer noted that she “received a fellow—a data guy who wanted to bait rats” using an algorithm developed in Chicago to prioritize locations for rat baiting. “Pittsburgh doesn’t bait rats. Stop trying to make rat baiting happen—that’s not what we need help on,” she said.¹²

2. DATA: Structured measurements about a subject

- Are the data attributes (variables) an agency collects on the subject, and the way this data is structured and stored, compatible with the model or use case?
- Is the data’s coverage (geographic, demographic, etc.) and granularity (level of detail) similar to those in the originating city?
- Is the data well-maintained and accurate?

Google Waze, a navigation app for smartphones, runs a data-sharing partnership with cities called the Connected Citizen program. In exchange for road closure information from municipal departments of transportation, the company provides urban planners with the product’s own data on slowdowns, collisions, roadkill, and other user-reported information. Waze data is formatted identically across municipalities, making methodologies for analyzing the data reusable across jurisdictions.

By contrast, while many jurisdictions capture data on restaurant inspections and health-code violations, the structure and quality of data are far from uniform. When Johns Hopkins’ GovEx attempted to replicate a predictive risk model for restaurant inspections from Chicago to other jurisdictions, it found that the data structure and quality varied so substantially that replicating the project ultimately required the development of a unique model.

3. MODEL: Abstract mathematical representation of phenomenon

- Is the analytics question the same across cities?
- Are the problem’s root causes known? Are they the same?
- Does the model represent similar behavior or phenomenon?

Many cities suffer the effects of abandoned or blighted properties. In some cities, vacancy is a condition of urban disinvestment and a relevant analytics use case is determining which properties to prioritize for demolition. In another city, landlords may hold vacant properties in speculation of higher sales prices, exacerbating an affordable housing crisis by artificially limiting the amount of available housing. In this case, the analytics question may be identifying candidates suitable for foreclosure prevention programs. Despite the fact that the symptom—residential vacancy—is the same, different underlying problems in different jurisdictions pose different questions and necessitate different approaches to analyzing data.

.....
¹² Laura Meixell (comments at Chief Data Officer Meeting, Cambridge, MA, September 24, 2015).

4. COMPUTATION: Automated procedure that transforms source data into information product

- Does the recipient city have adequate data processing resources to conduct the analysis?
- Are there legal or normative restrictions on what data can be processed, and how?

PredPol is one of the early vendors of commercial predictive policing software, which uses historical data on crimes to predict locations of future criminal activity. When PredPol was negotiating a contract with the City of Oakland, University of California graduate students launched a vigorous campaign against predictive policing in the city, claiming that predictive policing reinforces racially biased results of policing patrols. PredPol, which disputes that its software leads to disparate impacts among different racial groups, ultimately withdrew from Oakland, citing a lack of resources to engage in protracted political fights. While the Oakland police department identified an analytics use case developed in another jurisdiction that could be used locally, a local political movement, owing to concerns over certain automated procedures being used on certain types of datasets, enacted conditions that made the algorithm’s implementation infeasible.

5. IMPLEMENTATION: Process to take action based on analytical results

- Is the business process through which the service is delivered similar?
- Are the criteria for making decisions and procedures for taking action based on information compatible?
- Are the administrative structures and modes of governance of the organizations consistent?

Many municipal police forces in the United States use “early-warning” or “early-intervention” software to help departmental supervisors identify officers who may commit future misconduct. Police departments capture similar data elements on officer behavior; trends about this behavior can be reasonably generalized across cities, making an analytical solution a good candidate for replication (several companies have taken commercial early-intervention systems to market). In 2015, an applied research team at University of Chicago’s Center for Data Science and Public Policy (DSaPP) partnered with the Charlotte-Mecklenburg Police Department to develop an early-intervention system using machine learning to identify specific features predictive of misconduct.¹³ Results shows that the model increased true positives and reduced false positives. With this indication of success, the model was replicated the following year in the Metropolitan Nashville Police Department (MNPND).¹⁴ The initial implementation of the model in MNPND resulted in an outsized strain on departmental resources, requiring DSaPP and Metro-Nashville to adjust the model to result in more feasible workloads better adapted to local circumstances.

In this case, the model, which was designed to aid departmental supervisors’ decisions on how to deal with “at-risk” cops, was generalizable. However, the effect of a model’s rate of false positives (an officer flagged for risk of an adverse event who would not have gone on to commit one) varies substantially when the intervention is sending a warning letter to the officer and their supervisor, versus group counseling or training, versus reassignment.

¹³ Crystal Cody et al., “Building Better Early Intervention Systems,” *The Police Chief* 83 (August 2016): 20.

¹⁴ Hareem Naveed et al., “Human Lessons Learned Implementing Early Intervention Systems in Charlotte and Nashville,” University of Chicago Data Science for Social Good blog, March 29, 2018, <https://dssg.uchicago.edu/2018/03/29/human-lessons-learned-implementing-early-intervention-systems-in-charlotte-and-nashville/>.

6. MAINTENANCE: Viability of use over time

- Do city personnel have the skills to interpret the results and continually adjust a model after its initial calibration?
- Will the project continue to be funded by the city or external partners after its implementation?

When external organizations partnering with a city function as the implementation agent, they run the risk that there will be insufficient local buy-in to sustain the model’s use following the engagement’s conclusion. It is best practice in applied analytics to continually monitor results and tweak deliverables, meaning that a city requires both technological and human capital, outside of temporary external implementation support, to sustain the model beyond its initial calibration.

Danielle Fulmer, director of business analytics for South Bend, noted that projects her unit tends to find adaptable are those with simple data models.¹⁵ Projects that leverage the most advanced data science techniques may be more mathematically interesting, but their use is limited if analysts cannot easily manipulate the model or frontline staff cannot understand the logic behind a model’s outputs.

RECOMMENDATIONS FOR THE FIELD

Research for this paper uncovered uneven levels of optimism about replication of analytics projects in government. Many interviewees felt that transferring algorithms was a worthwhile cause, but missed more fruitful opportunities for inter-city collaboration. This was compounded by an apparent lack of consensus on the definition of “analytics,” which, in some cases, was taken to mean business intelligence software and data management tools and, in others, inferential statistical models. These gaps revealed areas in which additional research and programmatic support would enable a more productive agenda for urban analytics adaptation.

Recommendations for Urban Analytics Practitioners

1. **Start with internal use case adaptation and knowledge sharing:** Analytics leadership in cities should gauge opportunities for reusing algorithms and tools across

.....
¹⁵ Danielle Fulmer, phone interview with author, August 29, 2018.

divisions within their city, especially for datasets that are commonly used across different departments. Facilitating peer-to-peer replication within the organization can build familiarity with procedures for exchanging knowledge about analytics deliverables without the risk of copying projects from other cities. This will also facilitate organizational learning and induce demand for skills development and capacity building that advanced analytics projects—replicated or not—require.

2. **Avoid cosmetic analytics:** A project motivated by a desire to project an image of a city or agency’s technological savviness should send up a red flag. Political Scientists Charles Shipan and Craig Volden found that, among various motivating factors for policy transfer, a cosmetic orientation (“how can we *appear* to be the same?”) is likely to result in a short-lived project with limited impact.¹⁶

Recommendations for Networks of Analytics Practitioners

3. **Connect analytics communities of practice with domain-specific communities of practice:** Communities of practice in specific urban policy domain areas have existing, well-defined pathways for sharing best practices and transferring innovation. Connecting urban analytics professionals with communities of practice in transportation, economic development, policing, and other areas can create opportunities for infusing analytical approaches into existing pathways for innovation diffusion in domain-specific areas. According to Pittsburgh’s Laura Meixell, “It’s powerful to have people come in from the buildings side, ops side, and help work on an analytics project—the quicker you can find the front line, the better.”¹⁷
4. **Move from point-to-point replication to an open innovation model of collaboration:** One chief data officer suggested a more collaborative approach than one-off replication attempts, in which a network of leaders specify a shared agenda and

.....
¹⁶ The factors of policy replication, according to the authors, include learning from early adopters, economic competition among cities to attract or retain economic activity, coercion by state government, and imitation of a larger city. Charles Shipan & Craig Volden, “The Mechanisms of Policy Diffusion,” *American Journal of Political Science* 52 Iss. 4 (October 2008): 840–857.

¹⁷ Laura Meixell (comments at Civic Analytics Network Convening, Cambridge, MA, October 26, 2016).

jointly contribute rather than replicate a single use case from city to city. Replicating *processes* for collaborative problem-solving, rather than replicating specific tools or products, may be productive.¹⁸

5. **Discuss failure:** The social sciences are in the throes of what some have described as an epistemic crisis due to publication bias, which describes the incentive for sharing only surprising or interesting scholarly work. As a result, research activity represented in scholarly press is not likely to represent all studies and research activity that is undertaken.¹⁹ Similarly, epistemic networks and policy clearing-houses are configured to share and scale “best” practice. Discussing failure, and the conditions of an analytics project that led to it, can facilitate learning that is overlooked in discussions of “what works.”

Recommendations for Future Research

6. **Study local innovation ecosystems:** The biggest barrier to successfully adapting analytics use cases is implementing change to business processes. Future research should study the relationship between data analytics processes and mechanisms for implementing process and organizational change in government agencies. Understanding existing relationships between analytics-focused roles, such as a chief data officers and GIS leaders, and other innovation principals tasked with process improvement, such as a chief performance officers or chief innovation officers, can help develop guidelines for aligning strategies for performance improvement.
7. **Develop methods and templates for documenting cost savings or other value measures for analytics projects:** In a 2013 evaluation of 19 criminal risk-assessment methodologies, researchers found that “in most cases, validity had only been examined in one or two studies” and that “frequently, those investigations were

.....
¹⁸ Joy Bonaguro, “3 Reasons Why Replication in Government is a Flawed Goal,” *GovLoop*, February 15, 2018, <https://www.govloop.com/community/blog/3-reasons-why-replication-in-government-is-a-flawed-goal/>.

¹⁹ Andrew Gelman, “Ethics and Statistics: Honesty and Transparency are Not Enough,” *CHANCE* 30 Iss. 1 (2017): 37–39.

completed by the same people who developed the instrument.”²⁰ This is endemic of a wider lack of evaluation in urban analytics. Measuring the benefit of improvements to decision-making processes is methodologically challenging, and local officials often avoid evaluating the success of analytical tools. Future research should establish procedures for local officials to evaluate cost savings, or other metrics of public value, to determine the impact of analytics use cases. Quantifying value can create benchmarks and set expectations for the value of projects before replication attempts begin. It may also help cities better quantify the asset value of specific datasets.²¹

- 8. Understand how local analytics products can produce commercial value:** The output of the federal government’s 17 National Labs, which play a role in “translating basic science to innovation,” is recognized as a public good that can spur business growth.²² Some labs have formal technology transfer teams that encourage use of their intellectual property outside of government. Data analytics teams within cities may play a similar role. Future study on the extent to which the output of municipal data analytics—especially analytical applications that use open data—can be commercialized in local economies will help validate the public benefit of analytics development and use case adaptation between cities.

CONCLUSION

Replicating analytics projects from best-in-class or larger, more mature peer organizations is an important first step in developing a robust analytics practice. According to Evan Stubbs, an expert in implementing analytics at scale in large organizations in the private sector, successful organizations “recognize that business analytics is a journey, not a destination.” Further:

.....
²⁰ Sarah Desmarais & Jay Sing, “Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States,” Council of State Governments Justice Center, March 2013, <https://csgjusticecenter.org/wp-content/uploads/2014/07/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf>.

²¹ The question of whether and how government should monetize its data and create revenue streams is a substantial, under-researched subject of its own.

²² <https://www.energy.gov/national-laboratories>, accessed September 9, 2018.

While there's a large gap between a single example of best practice and consistent use across the entire organization, they understand that there are benefits to replicating automated methods. Most important, they take constant steps toward best practice, slowly making it pervasive across as many business processes as they can.²³

From this stage, organizations move from finite series of single “point” solutions to a structural capacity to adopt and scale analytics: “Their point of view shifts from one of ‘fix this, fix that’ to one of continuous improvement where best practices are identified, nurtured, and replicated across the entire organization.”²⁴

Several interviewees discussed the importance of inspiration, or the ability to expand the imagination to envision possible alternatives in their current operating environment. The existence of a use case in another city can help spur local stakeholders into action. But an analytics use case will remain a proof of concept if it is reproduced with little regard for the non-technical, cultural aspects of how unique bureaucratic and political organizations solve problems.

Whether solutions are transferred through peer networks, an agent of diffusion such as a university, or a company with an commercialized analytical tool, successfully connecting use case replication to a city's long-term analytics “journey” requires a deep investment in data and a sustained appetite for innovation—the foundational layers of problem-solving that cannot be copied.

.....
²³ Evan Stubbs, “The Intelligent Enterprise” in *Big Data, Big Innovation: Enabling Competitive Differentiation through Business Analytics* (Hoboken, NJ: Wiley, 2014): 90.

²⁴ *Ibid.*, 91.

CASE STUDIES

New Orleans Smoke Alarm Distribution

Local Problem

Following a high-profile fire that resulted in five deaths within one family in a home without a fire alarm, the New Orleans Fire Department searched for a more proactive process.²⁵

Analytics Question

Based on existing datasets, which properties likely not to have a smoke detector are at greatest risk of fire fatality?

Deliverable - Origin

Risk model, provided as a map and ranked list to guide a fire department in a door-to-door inspection task.

Replication Medium

“Smoke Signals,” a publicly available tool created by Enigma, a data-management and analytics company, and its publicly available code on GitHub. The tool allows all major American cities to upload their data and configure the model to their own jurisdictions.

Results of Replication

Model adapted and deployed for pre-pilot analysis in Syracuse, New York; model adapted and successfully deployed in South Bend, Indiana.

Background

Following its technical assistance to develop the smoke alarm distribution model in New Orleans, Enigma developed a tool that applied the same analysis to all major urban areas. Marc DaCosta, co-founder of Enigma, told Harvard’s Data-Smart Cities: “For us, one of the main goals throughout the process was to be extremely open and

.....
²⁵ Katherine Hillenbrand, “Predicting Fire Risk: From New Orleans to a Nationwide Tool,” *Data-Smart City Solutions* June 9, 2016, <https://datasmart.ash.harvard.edu/news/article/predicting-fire-risk-from-new-orleans-to-a-nationwide-tool-846>.

transparent in all of the components: the geocoder, the algorithm, and other tools, with the hope that anyone with an interest can go on GitHub and get them.”²⁶

Sam Edelstein, chief data officer for the City of Syracuse, knew that his city’s fire department had a similar program for distributing free smoke alarms. After discovering the tool and connecting with Enigma, he demoed the tool to Syracuse’s fire department to the excitement of departmental personnel. The results confirmed some patterns they had observed in their own work, which created a sense of trust in the technology. The tool also provided new, surprising insights on how smoke detectors were likely distributed in neighborhoods they thought they knew well. Edelstein said that the tool helped him show personnel “why it’s important for incident reporting to input data correctly—by doing this right, we can do things like [predictive analytics].”²⁷

In South Bend, the SB-Stat team, led by Director of Business Analytics Danielle Fulmer, works with agencies to set performance goals and provides business analysis and data analytics support to help achieve them. The fire department set a goal around distributing more smoke detectors through an existing community-outreach effort. Under the current practice, residents needed to submit a request to receive a smoke detector. Occasionally, the department would conduct “blitzes” in neighborhoods they knew lacked them. Fulmer discovered the Smoke Signals code on New Orleans’ GitHub page and used census data, along with local fire data, to reproduce the risk model. Her team worked with the fire department to create maps showing properties where the blitzes had already distributed smoke alarms. They found that the number of homes needing smoke detectors was higher than anticipated and higher than what was measured by the model. Fulmer speculates the 2011 Census data on which the model was built was stale. The project concluded in 2017 and was deemed by local stakeholders a successful replication effort, with room for improvement pending fresh data.

Key Takeaways

- *Publicly available analytic tools show the value of data analytics and the potential for “quick wins” to city departments. When a company like Enigma cleans*

.....
²⁶ Ibid.

²⁷ Sam Edelstein, phone interview with author, February 28, 2018.

data and develops a free tool, it lowers the perceived barrier to entry for analytics projects.

- *Successful analytics starts with collecting good data—an example of a predictive analytics can help make the case.* Convincing a department to invest in data collection when there is not a concrete use of the data can be challenging. Analytics use cases piloted by other cities can demonstrate the value of thorough data collection and management. According to Fulmer, the positive press that the South Bend Fire Department received around the project helped make the case to the department to invest in better data collection.
- *Replicating projects with publicly available data is more feasible than projects relying on custom data generated by departments.* Smoke Signals relied in part on census data, which is structured similarly for every municipality. Further, Smoke Signals did not involve advanced machine learning, making the process and results more interpretable to city staff.

Predictive Policing Software

Local Problem

In Los Angeles, when police officers are not answering calls for service, they are traveling routes at the discretion of individual patrol officers. This discretionary patrol is not explicitly informed by information on previously reported crimes.

Analytics Question

Based on past crime data, in what geographic segment of the city is crime likely to occur at a given moment and time?

Deliverable - Origin

Algorithm assigning a risk score to every unit in a 150 × 150-meter grid. Every box receives a risk score each hour; these results are aggregated at the shift level (e.g., day, swing, night). A fixed number of boxes within each patrol area with the highest risk are displayed for officers to patrol.

Replication Medium

Commercially monetized algorithm available through PredPol software.

Results of Replication

Algorithm used in 50 departments on three continents.

Background

When he was commissioner of the Los Angeles Police Department (LAPD) in the early 2000s, Bill Bratton wanted to replicate the CompStat methodology of data-informed decision making and performance improvement pioneered in the 1990s in New York City. Bratton asked mathematicians and criminologists at local research universities how the CompStat methodology could be extrapolated into the future. The goal was to prevent crimes with preemptive officer placement rather merely responding to crime with arrests and prosecution. This approach relies on a criminological theory that asserts that the presence of a “capable guardian” randomly and at multiple occasions in a given geography creates a deterrent effect for crime. The actual rollout and testing of predictive policing at LAPD was driven by then-Sergeant Sean Malinowski (now Deputy Chief of LAPD) and took place under the command of Chief Charlie Beck, about two years after Bratton’s departure from the department.

Researchers built a forecasting model with a risk score for certain geographies. After a crime analyst in Santa Cruz, California, read about indications of success in Los Angeles including a 30% drop for some crime types in a six-month period, he decided that predictive analytics would likely be a valuable tool in multiple jurisdictions. This analyst transformed the R code and past crimes data into map coordinates, worked with UCLA and Santa Clara University researchers, and developed PredPol, a company that marketed the software to different police departments. PredPol continues to work with Los Angeles as its main implementation partner, testing new methods and refining the product it markets elsewhere.

Key Takeaways

- *No cities are exactly the same.* “Each has very different philosophies of approaching the job that they do—that’s the nature of law enforcement,” said Brian MacDonald, PredPol CEO.²⁸
- *Simpler solutions are often more elegant and usable—and more effective.* MacDonald said that “it is possible to have too much data, and too much irrelevant data,” which can over-fit models. Limiting the PredPol model to only three variables—people, place, and time of past crime incidents—makes the product more applicable in multiple jurisdictions. Successful implementation depends on “ease of deployment and ease of use,” he said.
- *Empower people close to the operation.* Bringing the algorithm to the department from city leadership in the mayor’s office “does not work,” according to MacDonald. The project champions in the police department “want to be the ‘guy’ who thought of [the innovative solution].”

.....
²⁸ Brian MacDonald, phone interview with author, February 27, 2018.

APPENDIX A. INSPIRATION AND IMPLEMENTATION VEHICLES

Below are two *inspiration* methods (peer learning communities and expert clearing-houses) and two *implementation vehicles* (pro bono technical assistance and procurement) for analytics innovation. These are neither mutually exclusive nor exhaustive, but rather prevailing trends in the municipal analytics community.

Inspiration Method: Peer Learning Communities

Sociologist Etienne Wenger introduced the notion of “communities of practice” in 1991 to explain how peer groups become collaborative learning communities. These groups facilitate knowledge sharing among practitioners who do common work but may not encounter one another in the regular course of business. Building on this work, Wenger developed the “community of practice” based on research on scaling innovation across divisions in a single corporation, such as Chrysler. It has been adopted to help spread best practices in cross-jurisdictional governance, such as multinational grants managed by the World Bank.²⁹ Organizations such as the U.S. Conference of Mayors, the National League of Cities, and the Project on Municipal Innovation create networks of executive staff in cities to share best practices and collaborate on shared policy goals. Other organizations facilitate learning in narrower areas, such as the Rockefeller Institute’s 100 Resilient Cities initiative, which funds a “chief resilience officer” in 100 global municipalities to share access to “solutions, service providers, and partners.”

Examples:

- The MetroLabs Network is a national body comprising 35 unique local partnerships between universities and the cities that house them, focused on applying data to solve urban problems. By connecting partnerships in a larger network, the program aims to enable learning between academic and policy-making institutions.³⁰

.....
²⁹ Etienne Wenger, Richard McDermott, & William Snyder, *Cultivating Communities of Practice* (Cambridge, MA: Harvard Business Press, 2002).

³⁰ <https://metrolabnetwork.org/>, accessed February 18, 2018.

- The Civic Analytics Network, a peer group of chief data officers and analytics principals from over 20 U.S. cities, convenes biannually at the Harvard Kennedy School to collectively advance the field of municipal data and analytics.³¹

Inspiration Method: Clearinghouses

Clearinghouses rely on third-party, professional evaluators or credentialed thought leaders to assess the effectiveness of a program or policy and then use their influence to encourage its spread. Domain-specific policy clearinghouses exist in many aspects of municipal government, assembling studies and serving as brokers of “what works” in such fields as criminal justice and social service delivery. Awards programs provide another mode of recognizing early adopters of innovative programs. The Innovations in American Government Award, for instance, identifies top programs in the public sector and funds winning jurisdictions \$100,000 to replicate them. Several universities have recently begun capturing examples of success using data in urban governance. Because analytics in public administration is only recently gaining ground as its own discipline, outside of its subject-specific application in fields such as land-use planning or criminology, the field lacks the scientific rigor of peer-reviewed journal articles and formal evaluation protocols, though no shortage of grey literature exists.

Examples:

- Data-Smart City Solutions, a web publication run through the Harvard Kennedy School’s Ash Center for Democratic Governance and Innovation, captures success stories of data use and analytics in American cities. It includes a “Solutions Search,” a database of 200 examples of how cities use data to create value.³²
- The Pew and MacArthur foundations teamed up to create the “Results First Evidence Gateway” project, a database that aggregates domain-specific evaluations of evidence-based programs from such clearinghouses as Blueprints for Healthy Youth Development, Coalition for Evidence-Based Policy,

.....
³¹ “Civic Analytics Network: Helping Cities Unlock the Power of Data,” *Ash Center Communiqué* (Spring 2017), <https://ash.harvard.edu/news/civic-analytics-network-helping-cities-unlock-power-data>.

³² Data-Smart City Solutions, Solutions Search, <https://datasmart.ash.harvard.edu/civic-analytics-network/solutions-search>, accessed June 1, 2018.

U.S. Department of Justice’s CrimeSolutions.gov, U.S. Substance Abuse and Mental Health Services Administration’s National Registry of Evidence-based Programs and Practices Promising Practices Network, U.S. Department of Education’s What Works Clearinghouse, and What Works in Reentry Clearinghouse.

Implementation Vehicle: Pro Bono Technical Assistance

Several universities, non-profits, and companies provide cities with pro bono technical assistance to help them apply analytics to urban issues. These partnerships offer municipal agencies an opportunity to apply analytical tools developed in other cities or industries to urban challenges. Requiring no commitment of public funding other than staff time, they provide low-risk avenues to use analytics and experiment with advanced tools. Some organizations offer free services because they have a vested interest in a certain outcome, such as a philanthropic foundation looking to change a particular condition for a target demographic or a company hoping to showcase the applicability of a product in a public context. Diligence and care from both the city and partner are required to align the intent of pro bono assistance with local priorities.

Examples:

- IBM offers pro bono technical consulting and data science support to an array of public interest organizations, including local government.
- Under the auspices of the Civic Analytics Network, the University of Chicago provides data-science support to cities on such use cases as an early intervention system for adverse police interactions.
- The Center for Government Excellence at Johns Hopkins University—“GovEx”—is an implementation partner in Bloomberg Philanthropies’ What Works Cities program that offers help on analytics use cases, open-data program development, performance management, and data-management practices. Rather than replicating specific solutions, it focuses on problem-solving methods.³³

.....
³³ “We steer clear of one best practice for everything,” said Andrew Nicklin, director of data practices at GovEx. “We try to replicate methodologies and not specific use cases.” Phone interview with author, February 23, 2018.

Implementation Vehicle: Procurement

Cities often acquire assistance analyzing data through paid consulting contracts and software procurement. With funding, cities can drive competition for optimal outcomes by soliciting contract bids. Analytics procurement ranges from off-the-shelf predictive algorithms to self-service business intelligence tools to feasibility studies that leverage statistical methods. In theory, competition for contracts creates a market that forces innovation and improves products. However, procurement in government is notoriously cumbersome. Due to the uneven distribution of analytical maturity in local governments, public officials may struggle to articulate tool specifications that are both well scoped and not overly determinative of a solution, and may not have appropriate process controls and safeguards in place to ensure that commercial algorithms hold up to public expectations and norms around data privacy and bias.

Examples:

- COMPAS (the Correctional Offender Management Profiling for Alternative Sanctions) is an algorithm developed by Equivant (formerly Northpointe) that assesses the likelihood of a defendant in a criminal case committing a second offense. It is commercially available and licensed by many jurisdictions across the country, including every correctional system in the state of New York outside of New York City. The use of COMPAS in the criminal justice system has been a subject of controversy, stirring debate about definitions of “fairness” in automated systems and risk-assessment algorithms.³⁴
- CityMart is a platform and consultancy that helps cities frame problems and needs to expand the field of possibilities, rather than artificially limiting the array of solutions by over-specifying requirements in the request for proposals. The company is a trusted partner by several major initiatives in public sector innovation and innovation diffusion, such as What Works Cities, 100 Resilient Cities, and Nesta.

.....
³⁴ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, “Machine Bias,” ProPublica, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.



A publication of the

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

616-495-0557
www.ash.harvard.edu