

# Fact-Based Policy: How Do State and Local Governments Accomplish It?

Justine S. Hastings



## MISSION STATEMENT

The Hamilton Project seeks to advance America's promise of opportunity, prosperity, and growth.

We believe that today's increasingly competitive global economy demands public policy ideas commensurate with the challenges of the 21st Century. The Project's economic strategy reflects a judgment that long-term prosperity is best achieved by fostering economic growth and broad participation in that growth, by enhancing individual economic security, and by embracing a role for effective government in making needed public investments.

Our strategy calls for combining public investment, a secure social safety net, and fiscal discipline. In that framework, the Project puts forward innovative proposals from leading economic thinkers — based on credible evidence and experience, not ideology or doctrine — to introduce new and effective policy options into the national debate.

The Project is named after Alexander Hamilton, the nation's first Treasury Secretary, who laid the foundation for the modern American economy. Hamilton stood for sound fiscal policy, believed that broad-based opportunity for advancement would drive American economic growth, and recognized that “prudent aids and encouragements on the part of government” are necessary to enhance and guide market forces. The guiding principles of the Project remain consistent with these views.





# Fact-Based Policy: How Do State and Local Governments Accomplish It?

Justine Hastings

*Brown University and Research Improving People's Lives*

JANUARY 2019

This policy proposal is a proposal from the author(s). As emphasized in The Hamilton Project's original strategy paper, the Project was designed in part to provide a forum for leading thinkers across the nation to put forward innovative and potentially important economic policy ideas that share the Project's broad goals of promoting economic growth, broad-based participation in growth, and economic security. The author(s) are invited to express their own ideas in policy papers, whether or not the Project's staff or advisory council agrees with the specific proposals. This policy paper is offered in that spirit.

BROOKINGS

# Abstract

There is growing demand for a genuinely accountable government which, even with limited resources, delivers programs and policies with meaningful, measurable impact. Rapid advances in technology support the use of data and science in the private sector to develop insights about what people need, innovate products and policies to meet those needs, and then measure their success. Government has the potential to be similarly impactful, prompting recent federal and state calls for government to use a data-driven approach to produce efficient and effective policy solutions. But how can state and local governments use data and science to deliver improved results to their constituents? This article highlights the key challenges to creating and supporting fact-based policy at the state and local level, and draws solutions and lessons learned from an innovative and scalable partnership model developed with the state of Rhode Island.

# Table of Contents

ABSTRACT	2
INTRODUCTION	4
OVERVIEW	6
DEVELOPING DATA RESOURCES TO SUPPORT FACT-BASED POLICY	10
CREATING A DATA LAKE: DEVELOPING THE DATA RESOURCES FOR POLICY INSIGHTS	14
REFINING POLICY GOALS AND MEASURING PROGRESS	18
EXPANDING IMPACT THROUGH FACT-BASED POLICY	24
QUESTIONS AND CONCERNS	25
CONCLUSION	27
APPENDIX	28
AUTHOR AND ACKNOWLEDGMENTS	30
ENDNOTES	31
REFERENCES	32

---

# Introduction

---

In every sphere, from poverty alleviation to economic opportunity and education to health care, policymakers at the state and local levels tackle some of the toughest problems facing society. It is increasingly acknowledged that public policy needs to be effective, efficient, and evidence-based to make measurable progress; it needs to meet goals of improving economic opportunity equally and economically—it needs to deliver *value*.

Demographic changes make it increasingly clear that this value is much needed. For many state and local governments, unfunded pension liabilities loom large in the near future, threatening resources that are needed to provide education, infrastructure, safety, and social safety-net programs (Novy-Marx and Rauh 2009). Moreover, many believe that there is still a long way to go toward solving decades-old challenges. Haskins (2011) presents evidence that real, per person spending on antipoverty programs has increased substantially since 1970. Although this spending reduced material deprivation after accounting for program benefits, the fraction of households with market incomes below the poverty line increased. Recent research demonstrates that opportunity and economic mobility are elusive for many Americans who live in persistently poor and disadvantaged communities, particularly for minority males (Chetty and Hendren 2018). Opportunity gaps are present at birth, with significant inequalities in income and wealth acquisition persisting across generations (Chetty et al. 2018). These gaps continue to grow with the individual into college, the labor force, adult health, and longevity (Conti, Heckman, and Urzua 2010). Meanwhile, real spending on public education per student has more than doubled over the past several decades, but standardized measures of achievement in mathematics and reading have remained flat (Hanushek and Raymond 2006; National Center for Education Statistics [NCES] 2013).

While these persistent challenges may have many explanations, data and measurement could help untangle underlying causes, point to solutions, and measure future success. Policymakers increasingly seek public policy innovations that can deliver measurable, scalable results at lower cost.

Concurrently, advances in technology have allowed members of the business community to use comprehensive data and

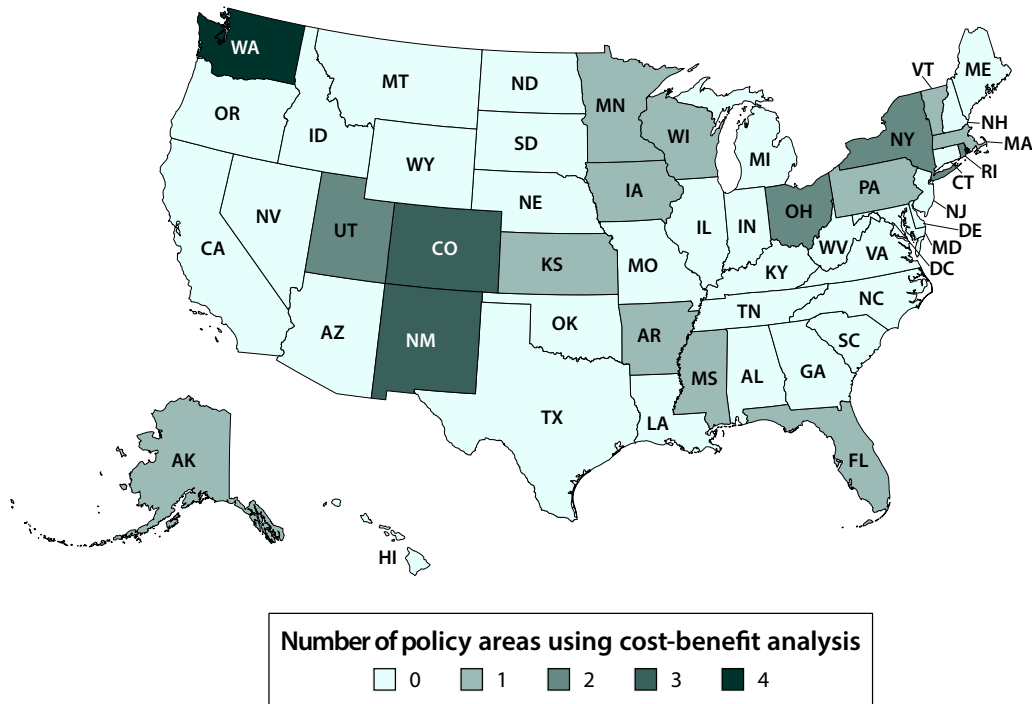
state-of-the-art technology to develop insights about what people need, innovate products and policies to meet those needs, and then measure the success of those products and policies. Many technology leaders take a strategic and quantitative approach; they have built data lakes of integrated data needed to produce insights quickly.<sup>1</sup> These leaders increasingly hire leading economists, statisticians, and data scientists from top research institutions to use data and science to measure how products work for people, and then develop new products that work even better (Athey and Luca 2018). They take a fact-based approach to product development. They ask, Does it save time? Does it better connect us to the people that matter to our business? Does it help us get our work done? Does it improve our overall quality of life? And does it deliver all this with value (e.g., at a price well below the benefits it confers)?

These are simple yet significant questions, and answering them scientifically on the basis of reliable data supports innovation that meets customer needs. Steeply falling prices of data and computing have made it possible to measure impact, and to design and deliver products in new ways that meet a wide range of needs at decreasing costs. Businesses and the private sector are not mandated to act or use these developments in the public interest. But what if these efficiencies in data resources and analytic approach could be harnessed for the public good? Government could use a similar framework to produce high-impact results for the communities it serves with efficient and effective public policy solutions.

There is a demand for a genuinely accountable government to deliver programs and policies that deliver meaningful, measurable improvements even in the face of increasingly limited resources. Government is beginning to respond to this demand. In 2015 President Obama issued executive orders to incorporate data and behavioral economics in policy design and evaluation during his second term in office (White House 2015). Senator Patty Murray and then-Speaker Paul Ryan established a bipartisan commission to investigate ways to promote data-driven policy (Ryan 2017–18). Murray and Ryan also produced a report that recommends creating data infrastructure that enables “a future in which rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy”

FIGURE 1.

## Cost-Benefit Analysis Use by State



Source: Pew-MacArthur Results First Initiative 2017; author's calculations.

Note: The Pew-MacArthur Results First Initiative's (2017) assessment of states' implementation of cost-benefit analysis covers four major issue areas: behavioral health, child welfare, criminal justice, and juvenile justice. States are counted as using cost-benefit analysis in a policy area if they conduct a report on the costs and monetized and/or nonmonetized benefits of multiple related programs. The analysis includes only examples that compare multiple programs within one analysis, excluding instances where states analyzed a single program. Data from the Pew-MacArthur Results First Initiative are available in the report's 2017 appendix.



(Commission on Evidence-Based Policymaking 2017). Several state and local policymakers have supported path-breaking work in developing data resources and put those resources to work in partnership with scientific research teams to help “state government operate at the speed of business” (RI.Gov 2018). Results for America (2018) is one such example of an organization that provides data useful for policymaking.

Not all states have made extensive use of data and evidence when making policy. The Pew-MacArthur Results First Initiative reviewed state governments' use of cost-benefit analysis across various policy areas, finding wide variation in state practices (Pew-MacArthur Results First Initiative 2017).<sup>2</sup> Figure 1 depicts this variation.

Of course, states might not conduct cost-benefit analysis because it is difficult to measure costs and benefits in the absence of readily available data, analytics, and a relevant body of scientific evaluations of policy alternatives. Reliable measures of policy benefits free from confounds and spurious correlations may be scarce, narrowly applicable, or unavailable. The dearth of facts with which to guide policy is caused in part by the lack of readily available data resources. Without the ability to use data quickly and reliably to produce the science needed to drive decisions, policymakers often have to make decisions based on arguments or anecdotes that are not supported by facts. They may invest in developing science

to inform their decisions, but without an optimized data resource and framework for collaboration, science may deliver results years too late to meet immediate decision needs.

Recent efforts to advance fact-based policy across the country have identified many of the challenges policymakers face in achieving the fact-based policy vision, indicating promising directions toward solutions. In this proposal, I review the key challenges faced by state and local policymakers, which include developing effective data resources, making those data resources useful for creating policy insights, refining policy goals and measuring progress, developing programs likely to be successful even in the face of countervailing market responses, and having the necessary technological resources and expertise to address these challenges.

I illustrate these challenges by drawing on examples from a research-policy partnership I started in Rhode Island in 2015, which is now a nonprofit network of scientists and faculty research affiliates called Research Improving People's Lives (RIPL, pronounced “ripple”). I draw comparisons to the practices that currently guide fact-based business policy. I highlight ways in which state policymakers can successfully move toward an institutionalized framework of fact-based policy, a framework that supports policymakers and nonprofit groups with data and science in their important work to develop and deliver policies and programs that improve lives.

---

# Overview

---

## FACT-BASED POLICY IN THE PRIVATE SECTOR

Private sector companies have thousands of products and programs that aim to serve customer needs in their careers and in their homes. Since the cost of computing has fallen precipitously, the private sector has quickly innovated by employing data insights to drive decisions and measure success. To do this, industry has increased investment in and use of data resources, research, and analytics. Unpacking the private sector data and insights development process is helpful for understanding how this process might similarly contribute to improvements in the public sector.

A recent Harvard Business School working paper by Athey and Luca (2018) highlights recent growth in private sector technology company efforts. The article discusses the technology sector model of building in-house insights labs by hiring and/or partnering with top academic researchers to build teams to analyze data and drive decisions.<sup>3</sup> Economists are often key team members because they have “a very particular set of skills [...] the ability to assess and interpret empirical relationships and work with data; the ability to understand and design markets and incentives, taking into account the information environment and strategic interactions; and the ability to understand industry structure and equilibrium behavior by firms” (Athey and Luca 2018, 5). Economists not only work with data, but they also work to achieve an “understanding [of] which relationships are causal—and which are not” (Athey and Luca 2018, 6), to avoid misinterpretation of data, and the “theoretical and empirical training [...] to think carefully about both intended and unintended consequences of different decisions” (Athey and Luca 2018, 8). These skills help businesses shape their business strategy so they can achieve their desired goals even when faced with broader market and policy responses (Athey and Luca 2018).

These teams of economists work with computer scientists, data scientists, and computer engineers to harness company data and turn them into timely insights using rigorous research and science. This process is fact-based policy for business. I use the term “fact-based policy” in this proposal to represent the union of data and the application of rigorous science to interpret the data in a way that accurately represents their context and measures likely outcomes. The term “data-

driven policy” may not be sufficient because data alone can be misinterpreted without proper expertise. For example, spurious correlations can be interpreted causally, leading to unreliable conclusions. By applying academic and statistical expertise to data, as Athey and Luca (2018) identify, assessing empirical and causal relationships, using empirical tools (e.g., instrumental variables, natural experiments, quasi-experiments, and randomized controlled trials [RCTs]), and incorporating incentives, selection, and market response into policy design—data can be appropriately interpreted to provide useful, empirically-grounded facts and forecasts for policy decisions.

## APPLYING THIS MODEL WITH STATE AND LOCAL GOVERNMENT

In 2015 I started a partnership with the Office of the Governor of Rhode Island to build and test a public sector equivalent of the technology sector model described above.<sup>4</sup> A key goal of Governor Gina Raimondo was to deliver better services, more opportunity, and economic growth to Rhode Islanders. Governor Raimondo assumed office in January 2015, inheriting a structural deficit of \$190 million (Vock 2015). At the time, Rhode Island was ranked 41<sup>st</sup> in the country for cash solvency, 48<sup>th</sup> for business environment, and 42<sup>nd</sup> for cost of living (CNBC 2015a, 2015b; Norcross and Gonzalez 2016). One of our top priorities was therefore to find ways to improve the efficiency and effectiveness of government programs.

I took a prototype approach, applying the private sector strategy described above to the problem of supporting fact-based public policy solutions. I built a scientific team to develop resources and ideas that could empower government partners to use data and science to drive decision making. We structured the organization to tackle the key challenges we foresaw in achieving the fact-based policy vision. Based on a history of research into policy-relevant questions in partnership with government (e.g., Beyer et al. 2015; Deming et al. 2016; Duarte and Hastings 2013; Hastings and Weinstein 2008; Hastings, Zimmerman, and Neilson 2015; Hastings et al. 2014; Hastings, Hortaçsu, and Syverson 2017) and the private sector (e.g., Gicheva et al. 2006; Hastings and Shapiro 2013, 2018; Hastings and Washington 2010; and Hastings, Kessler, and Shapiro 2018), in addition to the technology industry model outlined above, we identified several keys



**Table 1. Mission Statements for Rhode Island Policy Partners**

RI Agency	Mission
Executive Office of Health and Human Services	“Assure access to high quality and cost effective services that foster the health, safety, and independence of all Rhode Islanders.” <sup>a</sup>
Department of Labor and Training (DLT)	“The RI Department of Labor and Training provides workforce development, workforce security and workforce protection to the state’s workers, employers and citizens. Through federal and state funding, it offers employment services, educational services and economic opportunity to both individuals and employers. DLT also protects the workforce by enforcing labor laws, prevailing wage rates and workplace health and safety standards. And, the department provides temporary income support to unemployed and temporarily disabled workers.” <sup>b</sup>
Department of Human Services	<p>“DHS is an organization of opportunity, giving a lifeline with a full continuum of services to ensure:</p> <ul style="list-style-type: none"> <li>• Families are strong, productive, healthy, and independent.</li> <li>• Adults are healthy and reach their maximum potential.</li> <li>• Children are safe, healthy, ready to learn and reach their full potential.</li> <li>• Elders and individuals with disabilities receive a full continuum of services to enhance their quality of life.</li> <li>• Veterans are cared for and honored.”<sup>c</sup></li> </ul>
Department of Corrections	“Contribute to public safety by maintaining a balanced correctional system of institutional and community programs that provide a range of custodial options, supervision and rehabilitative services in order to facilitate successful offender reentry into the community upon release.” <sup>d</sup>
Department of Children, Youth and Families	“Partner with families and communities to raise safe and healthy children and youth in a caring environment.” <sup>e</sup>
Office of the Secretary of State	“The Rhode Island Department of State engages and empowers all Rhode Islanders by making government more accessible and transparent, encouraging civic pride, enhancing commerce and ensuring that elections are fair, fast and accurate.” <sup>f</sup>
Rhode Island Police Chief’s Association	“Assist its members – and all law enforcement officers in the State of Rhode Island – with the administration of public safety, to promote harmony and trust between law enforcement and the public, to enhance the effectiveness of law enforcement in the State, to strengthen public confidence in the police profession and to improve the quality of life in the communities we serve.” <sup>g</sup>
Department of Education	“Preparing every Rhode Island student for success in college, careers, and life.” <sup>h</sup>
Department of Health	“To prevent disease and protect and promote the health and safety of the people of Rhode Island.” <sup>i</sup>
Division of Public Utilities and Carriers	“The Division of Public Utilities and Carriers is a governmental body charged with the supervision and execution of all laws relating to public utilities and carriers and all regulations and orders of the Commission governing the conduct and charges of public utilities. These responsibilities include evaluating fitness and public convenience and necessity for motor, air, railway and water carrier services and competing providers of gas and electric service, fixing standards for utility service, witnessing the testing of measuring devices, ordering refunds to provide remedial relief, authorizing the issuance of securities, approving certain transactions between utilities, conducting investigations and holding hearings.” <sup>j</sup>

Source: a. Rhode Island Executive Office of Health and Human Services 2018; b. Rhode Island Department of Labor and Training n.d.; c. Rhode Island Department of Human Services 2019; d. Rhode Island Department of Corrections n.d.; e. Rhode Island Department of Children, Youth and Families 2016; f. Rhode Island Department of State 2019; g. Rhode Island Police Chief’s Association 2017; h. Rhode Island Department of Education 2019; i. Rhode Island Department of Health 2019; j. Rhode Island Public Utilities Commission and Division of Public Utilities and Carriers 2018.



to success: partnering with government leaders to identify and articulate policy problems; developing data resources that could be used quickly and reliably to generate facts that inform policy decisions; and developing fact-based solutions through collaboration between policymakers and scientists to ensure applicability, timeliness, and likely success.

Developing insights on a policy timeline requires data resources that are already well documented, easy to use, and accessible to research teams. This avoids lengthy delays which can result from building up new resources from scratch for each new project or purpose. The data resources also need to meet federal and state privacy laws and data security requirements. Thus, we followed the private sector model and assembled a team of computer scientists, economists, and policy analysts to collaborate with government partners. The team worked to first understand policy problems and then to generate the data resources to provide timely, reliable insights that could be used to develop effective policy.

The first step in achieving fact-based policy is the collaborative support of policy leaders who are committed to developing the resources and partnerships needed to use data and science to improve policy. Our work began by engaging with the governor and members of her cabinet to identify policy goals, and to understand the roadblocks that hamper implementation of fact-based policies. Leaders and agency team members across the Rhode Island government stepped forward to contribute.

We began by defining the impacts that leaders hope their programs and policies will have on the lives of all Rhode Islanders. It is often more difficult to articulate and measure objectives in public policy than in business. For example, a for-profit, publicly traded corporation has the objective of maximizing shareholder value (Lazonick and O'Sullivan 2000). Each policy the firm considers could therefore have a data-driven assessment of how that policy is expected to contribute to the objective of maximizing shareholder value. Assets and resources within the firm can then be shifted toward policies that demonstrate a larger return or contribution to that objective.

In government, each agency has a mission statement, which is a statement of its objectives. Table 1 lists the numerous multifaceted agency missions in the Rhode Island context.

We collaborated to understand key areas of need, and to understand policymakers' goals for working toward meeting their mission effectively and efficiently. Table 2 lists our interpretation of some of the policy goals and needs that leaders articulated when we first met to develop our collaborative, fact-based approach.

We reached several conclusions from the broad goals articulated in table 2. First, each goal is economically and socially important. That is, they all aim to alleviate poverty,

increase equal economic opportunity, and reduce program and administration costs. Second, many goals cross traditional agency boundaries. To measure policy success, integrated data from several agencies may be needed. For example, improving workforce quality may not only need labor data, but health, disability, crime, and incarceration data as well. Moreover, one agency's policies and goals may unintentionally impact the ability of other agencies to achieve their own goals. For example, a policy designed to reduce electricity costs may increase emergency service costs if electricity shut-offs lead to increases in emergency service use among impacted, low-income families. Third, data from outside Rhode Island government may be needed. For example, in order to develop programs to reduce food security, policymakers and research partners might need data from retail grocers or food pantries and soup kitchens to assess food purchases, food pricing, community demand for support, and fresh food availability relative to needs.<sup>5</sup>

Finally, progress toward achieving these goals will be difficult without the necessary data resources and expertise to use those data resources to guide policy direction. The ability to use the data to measure effectiveness of both current policy and policy innovations is crucial in this process.

Thus, we developed a fully-integrated approach, mirroring the private-sector model outlined above. Our model partnered scientists with engineers and policymakers. By combining their three complementary skillsets, we have the necessary components to create fact-based policy:

1. Scientists and engineers collaborate to develop the data resources needed to fuel insights from raw data, anonymously, securely, in compliance with privacy laws, and informed by policymaker experience.
2. Scientists and policymakers develop the list of insights needed to further agency missions. Scientists work with engineers to ensure the data lake is efficient and effective for a broad range of insight purposes, and that it delivers timely, reliable, and robust results.
3. Scientists and policymakers collaborate to deliver insights that both meet scientific standards and accurately reflect practical policy realities.
4. Engineers, scientists, and policymakers collaborate to develop best practices to ensure confidence, transparency, reliability, and scalability of the data resources and insights they produce.

In the next sections, I detail why I view each of these components as being important for fact-based policy, the key challenges to developing and delivering each of these components, and why this integrated approach helps address these challenges. I draw on my experience in Rhode Island to illustrate challenges and solutions.

TABLE 2.

## Examples of Policy Goals

Source of Goal	Policy Goal
Office of the Governor and Department of Corrections	Governor Gina Raimondo wants to reduce the three-year recidivism rate from 52 percent to 44 percent by 2020. What are some low-cost ways to meet this goal?
Office of the Governor and Department of Education, Division of Higher Education	Bachelor's degrees from high-quality institutions increase future earnings and help break the cycle of poverty. Rhode Island seeks to boost the rate of post-secondary degrees, especially at four-year institutions, among low-income students, with a low-cost but effective scholarship program
Office of the Governor, Children's Cabinet	In 2015, only 37 percent of Rhode Island's third graders scored proficient or higher in reading tests. Governor Gina Raimondo's goal is that "by 2025 [...] three out of four will be reading at grade level." How can early childhood programs help reach this goal?
Office of the Governor and Department of Labor and Training (DLT)	Governor Gina Raimondo's goal is to prepare Rhode Island's workers for the jobs of the future. Which skills are most in demand and how, if at all, is Rhode Island meeting that demand? What can we do to meet that demand?
Office of the Governor	Governor Gina Raimondo has set a target of reducing food insecurity in Rhode Island from the current rate of 12 percent to below 10 percent by 2020. Are there any low-cost solutions to meet this goal?
Office of the Governor and DLT	Governor Gina Raimondo intends to align job training programs with the labor needs of Rhode Island employers to ensure that Rhode Islanders have the skills to remain competitive in the workforce and in order to get people back to work and grow the state's economy.
Office of the Governor and Executive Office of Health and Human Services (EOHHS)	Rhode Island is currently one of the highest Medicaid-cost-per-capita states in the country. How can we lower costs while still providing quality care?
Department of Children, Youth and Families (DCYF) and Children's Cabinet	The Rhode Island DCYF wants to know if we are doing our best to help abused and neglected children. Should we be leaving children with families or removing them more often? What can we do to help troubled families improve their home environment to benefit their children?
Office of the Governor and EOHHS	Rhode Island is one of the states hardest hit by the opioid crisis. How can the Governor effectively combat the opioid crisis?
Division of Public Utilities and Carriers	Rhode Island electricity costs are some of the highest cost in the country. How can we ensure we help low-income families keep the lights on with simple, cheap, and effective policies which help them and lower overall costs of electricity administration?
Secretary of State	We passed a Photo ID law for voting. What is its impact on our efforts to ensure that are elections are fair, fast, and accurate?
Office of the Governor, State Police, Police Departments	Can we measure if policing is equitable, and helps reduce crime and supports working families? How can we support police with data and facts to improve policing policy?
DLT	How can we innovate to help every Rhode Islander succeed in their career?

---

# Step 1. Developing Data Resources to Support Fact-Based Policy

---

The first component necessary for producing fact-based policy is a data resource that can generate facts to guide policy and measure impact in a timely, robust, and reliable way.

The raw materials for such a data resource can often be found scattered through existing state agencies. Most government programs have administrative records that track a host of measures including, for example, applications, enrollments, services, and payments. However, those records typically are not in forms that are suitable for developing insights, are often housed and operated by third-party vendors, and are accessible only through graphical user interfaces for case management or to run preprogrammed reports for case management and administration. Even if raw back-end data were readily available, the data are often undocumented (no clear codebooks exist to explain the variables in the database and their meaning), or the vendor considers the documentation confidential information and is reluctant to share it with its customers. Thus, even if the raw data were available to policymakers, they would be in a form that is not readily usable for insights within an agency, let alone across agencies or in partnership with external experts.

We set out to develop such a data resource in Rhode Island. We uncovered and tackled several key challenges along the way. Below, I describe each challenge, and the solution needed to overcome it. I have categorized these into five challenges and solutions related to initial data resource construction.

***Challenge 1. Data exist in various formats and in layouts meant for supporting back-end administrative processes such as claims and applications, rather than for supporting insights. Backend data need to be re-organized to prepare them for transformation, cross-agency joins, and analysis.***

Agencies work with separate vendors with uniquely structured database systems. Some agency data may exist on legacy mainframe systems, custom-made systems, or in Oracle systems. Each agency is different (for example, human services databases are different from education databases), but many states use the same vendor, as there are scale economies in developing and servicing these back-end databases.

- Example 1. It is difficult to determine which individuals are receiving SNAP benefits at any point in time. The Rhode Island Department of Human Services' records for the SNAP program prior to 2016 consisted of versions of cases (multiple rows per case that record any activity on a case), and include multiple tables with hundreds of columns, many of which record internal functions that are important for case management but not for developing policy insights. Individual records linked to each case are stored in separate tables in JavaScript Object Notation (JSON) format, which makes them unusable for large-scale insights without transformation and processing. We conducted this key transformation to make these data usable for insights.
- Example 2. Most municipal police agencies in Rhode Island use a common vendor for their computer-aided dispatch (CAD) systems. The database is unwieldy for insights, and was developed for optimally supporting CAD operations. It contains more than 70 tables with thousands of columns. These columns often do not have primary keys (i.e., variables that uniquely identify records), which are important for combining and reshaping tables into data on important types of events, such as traffic stops or searches. Additional processing and inference are required to transform the underlying data into a relational structure that supports insights related to primary policing practices (e.g., calls for service and traffic stops). We produced correctly documented, derived tables, enabling insights for police policy that can be quickly and confidently developed.
- Example 3. Many agencies' small programs may not even have documented systems. For example, federally funded home visiting programs, with annual expenditures of \$400 million (Maternal and Child Health Bureau 2018), are relatively new programs. Documentation of enrollment and participation may be kept with dozens of NGOs who administer the programs but may have limited technology for recording system participation. It is often difficult to integrate the data of small programs, but these programs can still contribute to and benefit from integrated government data.

With the combination of computer engineering skills, scientific knowledge, and policy experience, we develop a

solution which curates back-end data and transforms them into understandable and reusable data fields for insights. For example, a table with each enrollee in the SNAP program in each month, size of household, number of children by age group, benefits given, key types of household income, and the rate at which benefits are spent down within the month can be very useful for any projects aimed at understanding the impact of the SNAP program on household health, education, or economic outcomes. By creating a table that distills the most important variables for research, we have optimized the data for insights in a reliable and sustainable way that can be used for future questions in a variety of policy spheres.

***Challenge 2. Data need to be clearly understood, defined, and documented so that reliable insights and research, can be produced and used to guide policy across projects and over time.***

Data across the public and private sectors often do not have proper documentation. Alternatively, vendors may choose to withhold proper documentation since such records may be considered a company trade secret (e.g., sharing it may enable the government agency to find alternative means to provide similar data services at lower costs). To use data effectively for insights, the data need to be documented so that variables are defined correctly, variable locations in the database can be identified easily, and government analysts and research partners can reliably use variables with the correct interpretation to draw the appropriate policy conclusions from their analyses.

- Example 1. Data from the Rhode Island Department of Children, Youth, and Families (DCYF) did not contain an identifier to match children to child protective investigators because program administrators did not need this link. However, the link is essential for a key DCYF policy question—measuring the impact of removal on child outcomes (Bald et al. 2019). We developed a method to match investigators to child cases, iterated with DCYF on this process, implemented it, and verified its accuracy with the DCYF team. Once these steps were completed, we documented the process and code to ensure the process could be reliably repeated for each update of the database. The data lake now contains the information needed to join these important pieces of information and facilitate future research and policy improvements.
- Example 2. The Department of Labor and Training (DLT) requires recipients of certain reemployment services to post their resumes to an online recruitment website called EmployRI. Data on EmployRI users' self-reported skills and job-searching activities (including job viewings and application submissions) can provide insights into the effectiveness of reemployment services. However, it is difficult to interpret these data because of a lack of

documentation from the website's vendor; additional exploratory analysis are required to determine the relational structure of the underlying database and the values corresponding to codes used in the database. We worked with the team at DLT to understand, verify, and create a codebook for these values; that codebook is now part of RIPL's database documentation.

The collaboration between engineers, scientists and policymakers facilitated the development of the codebooks and documentation needed to use the data reliably and confidently across a range of policy-relevant projects.

***Challenge 3. Sensitive data need to be anonymized, secured, and monitored to protect privacy and confidentiality. We developed a process to ensure security and anonymity in joined administrative records.***

Data need to be housed and analysis performed in an environment that meets security standards of federal privacy laws governing the data (e.g., U.S. Department of Justice 2018; FERPA 1974; FISMA 2002; HIPAA 1996). Broadly speaking, these laws aim to safeguard the security and privacy of individuals' records.

We recommend going above and beyond these requirements. In collaborative discussions, our team decided that a key principle for responsible data analysis is to keep the minimum amount of information necessary to accomplish the intended goals. As such, data should be anonymized, and personally identifiable information (PII) should be removed from the data to protect confidentiality, even if this is not required by a corresponding federal privacy law. For example, to generate research insights, personal identifiers such as name, address, and social security number are never needed. Robust research relies on anonymized data, and no one individual should drive results. Therefore, no individual identities are needed to generate data analytics and research insights.

Moreover, only secured access should be permitted, with automated logging and monitoring to ensure that only approved researchers have access to data. This ensures that the data are used only for intended purposes. Data should be analyzed on a central system that allows for automated and tamper-proof logs and auditing, and that does not allow individuals to download information from a system. Approval processes for offloading information from the system permanently documents any information downloads, and all downloads should be reviewed by a designated team and determined to comply with existing data use agreements for work tied to approved projects. Secure and cloud-based solutions now provide the most efficient and effective way to meet these standards, and can provide the government with direct access to auditing and logging while collaborating with nongovernment research teams. Hastings et al. (2018) outline

the securitization and anonymization processes we developed for our data lake.

***Challenge 4. Data are siloed within agencies, and need to be integrated to deliver policy insights, while preserving anonymity of records. Engineers and scientists collaborated to develop a process to accomplish this.***

As noted earlier, developing policies that can help improve a program or measure its success likely requires data to be combined from several agencies. At first glance, it might appear impossible to merge distinct datasets without compromising anonymity of individuals. However, these challenges can be surmounted. We developed a process for anonymizing and joining records across agencies that uses encryption and several layers of security controls (see Hastings et al. 2018). We designed a data import process that automatically encrypts PII and sensitive identifiers, allowing records to be joined across agencies based instead on a global anonymized identifier. PII is used only to construct the anonymized identifier, and is never accessible for analysis. Thus, important insights can be developed cross-agency by allowing researchers with approved access to join tables across agency domains using the anonymized identifier only.

- Example 1. DCYF asked RIPL to explore how data and science can be used to identify at-risk families to target proactive support services before a situation becomes acute and causes costly investigations and challenging removal decisions. This required anonymized birth records alongside household relationship records from the Rhode Island Department of Human Services to form the underlying population of children and families for the model; top predictors in the model came from a wide range of data sources, including Medicaid claims from the Rhode Island Executive Office of Health and Human Services (EOHHS), social program utilization and household composition from the state's Department of Human Services, and incarceration history from the Rhode

Island Department of Corrections. Some of the strongest predictors came from data outside the agency (e.g., crime and health records; see EOHHS 2018).

- Example 2. Measuring the return on early childhood interventions through increased educational attainment and reduced social service costs later in childhood is important for making wise policy investments. Estimating program impact requires data on births, educational outcomes such as standardized test scores, and social program enrollment to be joined (anonymously) to connect baseline characteristics with program receipt, and to measure impact on key metrics of development and achievement.

***Challenge 5. Federal data privacy laws require that data be accessible only for approved projects whose objectives are in line with agency missions. Projects need to be defined and tied to specific data tables, and access restrictions must be implemented and documented so that only the data needed for each approved project is accessible by project researchers and analysts.***

To comply with federal privacy laws, data must be made available only for approved projects. This is because privacy laws typically require nonagency use of data (both within-government nonagency data use, and external-to-government nonagency data use) to further the mission of the agency (see, e.g., FERPA 1974; HIPAA 1996; and U.S. Department of Justice 2018). Here again, the fully integrated model allowed us to develop the data resources needed to deliver commensurately defined insights that further the missions of the contributing agencies. The scientists, engineers, and policymakers are all key collaborators throughout the data design, build, ideation, and insights development processes.

Box 1 summarizes this paper's proposals that pertain to the initial development of state data resources.

BOX 1.

## Policy Proposals for Developing Data Resources to Support Fact-Based Policy

- Modern secure cloud computing allows government to keep control of its data with full transparency into data access and use, and enables research partnerships to develop insights with efficiency and effectiveness at the highest levels of security.
- Computer scientists, economists, engineers, and policy experts should work together to reorganize and prepare data from administrative databases to optimize the database for policy insights.
- To use data effectively, the data need to be documented so that variables are defined correctly, variable locations in the database can be identified easily, and government analysts and research partners can correctly interpret variables and thereby draw the appropriate policy conclusions from their analyses.
- Data need to be anonymized, housed, and used in ways that protect confidentiality and exceed the security standards established in federal privacy laws.
- Disconnected datasets in agency silos should be encrypted and merged using anonymized identifiers that meet privacy requirements and safeguard the security of individuals' records.
- Data should only be made available for approved projects that advance the mission of the relevant state agency.

---

## Step 2. Creating a Data Lake: Developing the Data Resources for Policy Insights

---

Throughout state government, data are recorded for program administration purposes, and not to generate insights. Building data tables for research without understanding research needs can lead to ineffective or underutilized data resources. Researchers and computer scientists need to set a goal together, keeping eventual research needs in mind, to define key data tables that are curated to contain reliable information that can be utilized for many purposes. Below I describe four challenges and solutions related to further developing data resources to optimize them for insights.

To enable fact-based policy, a data resource must be built, documented, optimized for insights, and readily accessible to government analysts and scientific teams. This is the standard within many private sector companies. Companies build derived tables, which aggregate or transform underlying transaction records into useful formats for scientific and policy teams. Derived tables improve efficiency and reduce the potential for errors because coding does not have to be repeatedly implemented for each user and each use. For example, in the retail sector, a table of customer monthly purchases may be created from underlying information about individual transactions to generate measures of total sales to each customer. Customer locations may be geocoded, additional external information can be added (e.g., credit rating, vehicle registration, or home ownership), and predictive analytics can be employed to categorize customers by key attributes. This helps ensure that heterogeneous customer needs are met across all customer segments. These data transformations are performed as part of an automated script, which is run to consistently generate data for insights development. The analogy for public policy might be a table with monthly program enrollment and benefits across all social programs, expenditures, and revenue streams (e.g., Supplemental Nutrition Assistance Program [SNAP], Temporary Assistance for Needy Families [TANF], and the Child Care Assistance Program, disability insurance, incarceration, education, and labor force participation), and a risk score for vulnerability to poverty or food insecurity, respectively. These tables and risk measures are reproduced by the same documented process each time the database is updated, so that scientists and analysts can rely on the data for projects across time and across policy domains.

*Challenge 6. Raw data can be cumbersome and time-consuming to use. It is useful to build derived tables that transform raw data into the data fields most commonly needed across projects to deliver policy and research insights. As a solution, we built the RI 360 data lake, which is a relational database with expertly constructed derived tables benefiting from scientific and policy-practitioner input.*

Our scientific and engineering team collaborated with policymakers to develop tables for high-level insights by transforming all back-end, raw data into usable data for measuring policy impact. To create these tables, we relied on expertise from scientists and policymakers to determine frequently-used variables from each program. We relied on data engineers to create automated scripts to reliably and reproducibly generate derived tables for these variables. The collection of derived tables was continually improved as new analysis needs were identified, improving the coverage and the usability of the data lake. We also constructed an RI 360 summary table to combine variables that are common to many analysis projects into a single, easy-to-use table. New projects can derive most of the variables they need from the summary table.

- Example 1. Incarceration impacts both incarcerated individuals and their families. Policymakers are interested in measuring the impact of parental incarceration on children's outcomes in school, as well as on their labor market outcomes. Controlling for prior incarceration can be important when measuring a program's impact on labor market outcomes. Thus, a useful table for many policy questions might list each anonymized individual and indicate for each month and year whether or not they were incarcerated, where they served their sentence, and on what category of charges. Although this sounds like a simple query, the incarceration system and its administrative back end have complex relationships between charges, sentences, incarceration, and parole. Determining from these databases, which are designed for prison operation, when an anonymized individual is incarcerated or released requires complex data processing. We created a derived table for incarceration spells, which lists all year-months in which an anonymized individual is incarcerated with a summary of the charges and sentences related to that



incarceration, allowing scientists and policymakers to instantaneously run regressions and data visualizations relating incarceration to family, labor, and health outcomes.

- Example 2. To aid researchers with joining these kinds of cross-agency data, we combined all of the derived tables into a single RI 360 summary table that spans 20 years of history for the state’s most important programs and outcomes, as well as demographic information about anonymized individuals (e.g. age, race, ethnicity, and gender). This table is a single source of information that on its own can provide the majority of variables needed for a new research project, exploratory analysis, or visualization. Like the derived tables, it is indexed by the anonymous unique individual identifier and year-month. Creating this single derived table ensures that all research in the lab draws from common variable construction and definitions—a process that is robust and reproducible. This also ensures that any changes based on new information are automatically pulled into all analyses in the lab. Thus, for example, many projects may use an indicator of whether a parent is on SNAP as an explanatory variable for a child’s health or education outcomes. Researchers can use the child’s anonymous identifier to pull an indicator from the RI 360 summary table as to whether that child had an incarcerated parent or was a member of a household who received SNAP benefits in a given month. All regression analysis pulls from the same table, thus ensuring that all analysis includes the same definition of family relationships and SNAP enrollment.

***Challenge 7. It can be complicated to define data fields, and that complication can lead to unspecified or misunderstood changes within and across projects if data construction is not documented and automated. Automating and documenting data ingest and construction processes makes the data lake and therefore the insights derived more reliable across projects and over time.***

Since the quality of documentation received from agencies can vary considerably, RIPL built an automated system to generate codebooks in a standardized format for all tables in the RI 360 data lake. A codebook is a reference document that describes each column in a table along with summary information about the type and range of values and the proportion of missing values. Codebooks help researchers identify data quality issues, and enable faster onboarding of new team members. We further integrated these codebooks into the wiki-based RIPL Knowledge Base so that researchers can add further notes and document new discoveries as they work with the data. This knowledge is thus preserved and shared throughout all projects to increase reliability at speed.

Generating proper documentation is more difficult than it might first appear: it often necessitates more than simply

describing existing data. For example, when working with the Rhode Island DCYF, solving this research challenge was an exercise in collaboration. Our team worked with DCYF to understand how data were physically entered into the DCYF reporting system, including shadowing their 24-hour maltreatment hotline to map the data entry process to back-end variables—i.e., the raw data that must be processed for use in research—in order to optimize the program administration data into derived fields usable for research. Once this investment has been undertaken for a dataset, notes and decisions are documented in the code and in the knowledge base to reproduce the data with each data lake update, and to preserve and share knowledge across projects, across team members, and over time.

***Challenge 8. Teams often benefit from insights that other researchers have generated. These benefits can be lost without standardization of knowledge and documentation across projects. To maximize efficiency and improve knowledge across sectors and across projects, and to ensure knowledge feeds back reliably into data improvements, standards for developing insights should be implemented as part of the process for using the data resource.***

The data lake is a living entity that improves and grows with use over many projects. It can produce insights that make use of all relevant information and are documented and reproducible. Every analysis conducted by RIPL is tied to a fixed version of the RI 360 database, and can be rerun against the research version at a later time to reproduce the results to help new teams learn, to resolve future questions, and to support robustness checks that may arise when insights lead to policy innovations to be tested in the field. To encourage reproducibility, analysis projects use a common project template to organize code and research results in a standardized way (e.g., RIPL 2018b). The common template enables reuse of methods between projects, ensuring scalability of knowledge developed. Finally, RIPL developed coding style guidelines to improve the readability of analysis code. Using consistent coding style reduces the cost of reviewing code, increases the reusability of code across projects, and lowers the start-up cost of new team members joining a project.

***Challenge 9. Projects involve multiple parties over time, requiring documentation and project management to ensure insights are developed efficiently, effectively, and securely.***

Several project management challenges arise when defining projects and developing insights. First, as mentioned earlier, projects need to further key agency missions in order to gain access to agency data. Second, knowledge may be developed during the research process and lost along the way if undocumented. When agency objectives change or team members change, the insights and analysis process should

## BOX 2.

# Lessons from RIPL

Although data are ubiquitous within and across government agencies, raw data are often not in a usable format for generating data insights. Data resources need to be built to support fact-based policy, but in a way that complies with federal and state privacy laws, and supports best practices (which may go above and beyond legal compliance) to ensure that policy insights are driven by robust and reliable research methods.

To do this requires collaboration between practitioners and expert scientists. Practitioners often are the key resource to know the ins and outs of data fields when underlying data are poorly documented. In the absence of code-books, individual practitioner knowledge is often needed to understand how to translate individual back-end data system fields into usable data formats for insights. Expert researchers are needed to determine what data formats are necessary for research insights; to understand what the right questions are to ask practitioners to develop a robust and reliable insights database; and to construct the technical methodology to transform siloed back-end data across diverse agencies and departments into an integrated system. Researchers must accomplish these tasks in compliance with federal and state privacy laws and best practices for research using sensitive data.

For industry, such research-practitioner collaborations may often happen in house because industry hires top computer scientists and economists to work with business leaders to drive fact-based strategies and business policy at the speed of industry and commerce. While this may not be feasible for government due to budget constraints and inherent hurdles in hiring and building agile teams, partnerships with top researchers are often possible, increasing objectivity in policy evaluation.

Our team built the initial RI 360 data lake with government partners in approximately 18 months with a budget of approximately \$2 million for producing both the database and make measurable progress toward policy and research insights.<sup>6</sup> Because data layouts are often the same across states for many major social programs, the marginal cost of producing the same database for additional states is lower because many of the same code, pipeline, and principles can be applied.

This pilot model produced the necessary resource for fact-based policy efficiently and in a time frame that could produce results and returns within one gubernatorial term. Together, our team used data to improve policy in education, health, crime, labor training, and child protection. We brought policy leaders, scientists, and community groups together around data and scientific results, enabling the state to undertake comprehensive interagency policy approaches not previously possible.

be well documented to ensure objectivity and to maximize insights. To solve these challenges, we did the following:

- Developed a formal chartering process with government partners to clearly define the project goal(s), the data fields needed, the key responsible points of contact in the agencies and scientific teams, the deliverables, and the overall timeline. These steps tied project development and policy goals to database field development and vice versa, creating a feedback loop that improves the data resource, the policy questions we seek to answer, and the insights we develop.
- Documented charters, along with all project progress in task management software that preserves and archives documents, decisions, tasks, and task completion with

visibility to all project members. Again, documented information fosters constant improvement in both data resources and in the generation of insights and knowledge.

- Documented analysis and coding decisions in task management software. Changes to code are versioned, archived, and documented using standard best practices for software development, so that decisions and changes in results can be traced, and new projects can learn from and build on top of past projects.<sup>7</sup>

Commitment to documentation, communication, and transparency through these processes allows scientific and policy teams to work together with increased confidence, and to produce reliable insights with data resources with a higher likelihood of translating into policy improvement. At the

same time, these processes feed back into improved database development, which in turn feeds into improved policy goal development and insights over time. Some of the lessons learned from our experience with RIPL are described in box 2.

Box 3 outlines the key components of my proposal to develop state data resources into a data lake—a resource with broad usefulness to policymakers and policy stakeholders.

I next discuss key challenges in insights development and policy impact, and how the science and policy teams worked together to tackle these challenges.

BOX 3.

## Policy Proposals for Turning Data Resources into a Data Lake that Yields Policy Insights

- Transform all back-end, raw data into useable data for measuring policy impact by creating a relational database with expertly constructed derived tables.
- Automate and document data ingest and construction processes to make the data lake—and therefore the insights derived—more reliable across projects and over time.
- Implement standards for research and insights in order to enhance robustness, reproducibility, and scalability.
- Develop a formal chartering process with government partners to clearly define the project goals, the data fields needed, the key points of contact in the agencies and scientific teams, the deliverables, and the deliverables timeline.

---

## Step 3. Refining Policy Goals and Measuring Progress

---

Over the course of the project, ideas and insights develop in tandem with data resources to support policy and research. The goal is to use data and science collaboratively with input from policymakers and scientific teams to meet important agency goals and thus improve lives through more efficient and effective policy.

A key first step in this process is refining broad policy goals into actionable next steps with answerable questions. Each of the policy goals and questions in table 2 are important and are clearly aligned with the missions of their respective agencies as outlined in table 1. However, each goal or question is very broad, and does not point to clear next steps for answering the question or achieving the goal. Given broad policy goals, it is unclear how to specifically apply facts and evidence to identify solutions, even with a data resource like the RI 360 data lake.

An integrated approach helps strategically employ the data lake to deliver robust and reliable insights on a policy timeline. Collaboration between the scientists, engineers, and policymakers to build the data lake translates quickly and seamlessly into the complementary production of insights, which helps to continuously improve the knowledge base and data resource. In this section, I highlight the key challenges in moving from the data to fact-based insights, and highlight solutions we developed as part of our integrated model.

***Challenge 10. Policy goals are broad and ambitious. Collaboration between scientists and policymakers, supported by an optimized data resource, can help distill these goals into clear, answerable and implementable next steps.***

Broad goals and questions require policymakers to define specifically what it means to achieve the goal or answer the question. This process has three key components. First, to achieve a goal and set out a path toward achievement, agencies must assess how close they are to achieving the goal in question. This assessment requires measurement and specificity. Second, a path needs to be developed from the current state to the desired outcome. This also requires measurement to quantify and track progress. Third, impact from policy innovation needs to be measured to assess if desired goals are being achieved.

Refining a broad policy goal into a specific, measurable objective often requires collaboration between policymakers and expert researchers to define measurable outcomes that reflect the goal's broad intent. Policymakers and policy analysts provide important local context, constraints, and needs. Expert researchers bring knowledge of existing evidence and research, a framework for evaluating which evidence is reliable, and a systematic approach for understanding relevant incentives and market forces. Finally, they provide technical expertise necessary for generating insights that refine the broad goal into actionable next steps.

- Example 1. Reducing recidivism: The governor has a goal of reducing three-year recidivism rates from the current 52 percent to 44 percent by the year 2020. The goal is important, but as stated does not provide clear directions for next steps. Next steps can be established once we identify which policies and programs, if any, can have a large impact (e.g., several percentage points) on reducing recidivism and which are achievable given a limited budget. RIPL researchers took the dual approach of (1) consolidating the existing research literature to identify low-cost policies or programs with promise for significantly reducing recidivism, and (2) using the RI 360 data lake and machine learning algorithms to measure which factors predict recidivism by searching over a broad range of factors to uncover the strongest predictors and assessing which predictors point to high-impact and low-cost policy solutions. Predictive models showed that particular in-prison training programs and connections with social programs upon release from prison impact recidivism rates by several percentage points each, and point to potential low-cost policy innovations.

Policymakers and scientists, who had access to an optimized data lake, quickly refined the broad policy question to these specifics: (1) Can we measure which in-prison programs are effective at reducing recidivism and shift enrollment to those programs to significantly reduce recidivism? And (2) Can we introduce programs to increase the rate at which prison releasees enroll in social safety-net programs as a low-cost way to reduce recidivism? Both of these questions are data driven, and point to clear next steps toward fact-based policy improvements.

- Example 2. Reducing Medicaid costs: Medicaid inflation-adjusted spending grew 63 percent from \$263 billion to \$429 billion from 2000 to 2012, with enrollments expanding 50 percent from 44 million in 2000 to 66 million in 2010 (Mann 2014). Rhode Island policymakers were interested in finding ways to reduce emergency department (ED) costs without compromising the current quality of health care that Rhode Islanders received. We began by consolidating the existing literature to assess the current evidence surrounding ED costs in the United States and understand if any programs currently exist to reduce costs. We found that Medicaid enrollees use the ED disproportionately more than privately insured individuals (Mann 2013). ED visits are a relatively expensive form of care, and have sharply increased in recent years. There is a broad consensus that many visits to the ED could be prevented by treatment in primary care (Billings 2013). Then, we used machine learning and the RI 360 data lake to measure which factors predict which patients in the bottom 50 percent of Medicaid costs will be in the upper 10 percent of costs in the coming year. In other words, we proactively identified need by predicting preventable high-cost (or cost-emergent) users before their costs increased. We identified four types of preventable ED costs, and produced a cost-benefit analysis for the state that compared the use of traditional wait-and-see methods of identifying high-cost users to the use of the predictive model to proactively identify need.

Here, the broad policy question was honed to this: Which current social assistance programs can be used as a low-cost point of outreach and connection for individuals predicted to become high-cost users of divertible ED care? This question points to a clear field trial that can be developed and tested in partnership with one of many nonprofit organizations that provide support for designing and implementing field trials. Results can be analyzed quickly with the data lake.

***Challenge 11. Data alone can be misleading. Therefore, it can be challenging to develop effective and efficient policy solutions without collaborating with scientists. Disentangling cause and effect, as well as competing explanations, may require collaboration between policymakers and scientists with access to an optimized data resource.***

After developing a data resource such as the RI 360 data lake and honing broad policy goals into data-driven policy paths, specific solutions must be developed to advance policy goals. Having the right data resource and a combination of scientific and practitioner expertise are necessary for producing timely evaluations, innovations, and fact-based policy decisions.

Partnerships between policymakers and scientists are critical to tackling this challenge. Scientists can bring expertise in causal analysis and predictive modeling, while policymakers

ensure programs are correctly interpreted and that insights are relevant to driving policy solutions.

- Example 1. Disentangling cause and effect in reducing recidivism through better use of in-prison training programs.

Following the earlier example of recidivism reduction, a clear next step is to determine which prison training programs lower recidivism the most. To make this determination, we cannot simply measure the mean recidivism of each program's enrollees. Prisoners with different backgrounds choose different programs and have different probabilities of returning to prison regardless of training. We do not want to label a program as effective at lowering recidivism if it simply attracts those individuals who are less likely to recidivate in the first place. Instead we want to measure value-added—how much a program reduces the probability of recidivism given an inmate's background and baseline probability of recidivating. In other words, we want to increase enrollment in programs that move the needle for their enrollees, reducing enrollees' likelihood of returning to prison. Rewarding programs that simply serve low-recidivism enrollees may be counterproductive in the short run and in the long run: programs working hard to reduce recidivism for those likely to return to prison are not recognized for their hard work, and are unfairly classified based on the difficulty of the problem they are trying to solve instead of being rewarded for progress.

Identifying these programs required machine learning to estimate the predicted probability of recidivism for each inmate, given their pre-incarceration background, then using regression analysis to determine how much each program lowers recidivism rates for its enrollees relative to their predicted baseline recidivism rates. The resulting analysis suggested that the most effective programs are those that provide job-specific skills in industries like construction that are likelier to hire former inmates, and that provide generally important basic skills and education. In contrast, programs with low mean recidivism rates do not appear to reduce recidivism, but instead tend to enroll inmates with low recidivism probability to begin with. For example, training programs providing college-equivalent courses in history and humanities have low recidivism rates. However, this is simply because they tend to enroll prisoners with low baseline likelihood of recidivating. Developing evaluation or performance measures for training programs without conditioning on baseline risk of enrollees can lead to unintended policy consequences (e.g. accidentally rewarding programs that enroll low-risk individuals instead of rewarding programs that lower recidivism risk) (RIPL 2017d).

Collaboration between scientists and policymakers can ensure that policy innovations meet their intended goals. Access to an optimized data resource ensures that the underlying facts can be discovered rapidly and reliably, and innovations can be efficiently and effectively evaluated once implemented.

- Example 2. Using data and science to identify and develop a low-cost program with likely success—the Rhode2College.

In 2018 Governor Raimondo set an ambitious goal to have 70 percent postsecondary achievement in Rhode Island by the year 2025, up from 47 percent in February 2018 (RI Gov 2018). College attendance is powerfully associated with improved long-term outcomes such as social mobility, economic well-being, health, and longevity (Autor 2014; Avery and Turner 2012; Heckman, Humphries, and Veramendi 2018; Oreopoulos and Salvanes 2011). But despite these benefits the national immediate college enrollment rate for low-income students is 20 percentage points lower than it is for high-income students (NCES 2017). This pattern is present in Rhode Island as well.

RIPL analyzed data from the Rhode Island Department of Education (RIDE) and found that from 2011 to 2014, the four-year college-going rate for Rhode Island high school graduates was 22 percent among students qualifying for free and reduced price lunch (FRL), but 51.5 percent for non-FRL students, an enrollment gap that persists even when adjusting for different levels of academic achievement between the groups (RIPL 2017a). Even when low- and high-income students have the same academic achievement, low-income students tend to enroll at colleges with lower returns on investment.

Interestingly, measures of college readiness themselves do not necessarily predict the college enrollment gap: while significantly fewer FRL students historically took the PSAT, predicted PSAT performance using machine learning models and 8th-grade standardized test scores (New England Common Assessment Program) suggests that those who historically did not take the PSAT likely would have performed similarly to those students who did take it. This suggests that many low-income students have the potential to be college ready and to enroll in college, but do not take the needed steps during high school to be college ready and to enroll after graduation. Such an early-life decision may close many doors to opportunity for these individuals throughout their lives.

This data-driven insight is consistent with research which shows that programs that support students—financially and otherwise—at all steps in the college-going process are the most effective in improving college attainment for low-income populations (Deming and Dynarski 2010). These

types of programs, such as providing individualized FAFSA filing assistance, can have results comparable to large-scale and expensive scholarship programs at a fraction of the cost (Bettinger et al. 2012). Programs that provide low-income students and parents with information about high-performing schools, the college application process, and financial aid for college have also been shown to increase test scores, improve college enrollment rates, and help students enroll in colleges that better match their academic aptitude (Bettinger et al. 2012; Hastings and Weinstein 2008; Hastings, Zimmerman, and Neilson 2015; Hoxby and Turner 2013; Jensen 2010; Thaler 2016).

Scientists and policymakers collaborated around the data-driven facts and prior research findings to design a low-cost, light-touch solution that addresses information barriers while providing short- and long-term financial incentives to low-income students. We partnered with RIDE, the Rhode Island Office of the Governor, and the College Board to implement Rhode2College, a step-by-step, achievement-focused program designed to give high school students the tools they need to succeed on the path to college. Rhode2College uses a texting platform that rewards students for treating the college application process as a marathon and not a sprint, by breaking down the process of getting to college from a large, daunting task into smaller discrete, achievable milestones. Students are rewarded each time they complete a milestone, and can earn up to \$2,000 while completing college application milestones. Drawing on a wealth of behavioral economic research, Rhode2College encourages both short- and long-term progress toward college goals by giving some money to students immediately, and putting some in a savings account they can access upon enrolling in college. Immediate and long-term rewards help students stay on track for the long haul. Importantly to the state, Rhode2College both functions as a scholarship and empowers students to take their future into their own hands. To date, Rhode2College has a 61 percent enrollment rate and 83 percent average milestone completion rate among enrollees. Minority and female students were more, instead of less, likely to take up the program (RIPL 2019).

- Example 3. Decreasing food insecurity in Rhode Island: SNAP Split-Issuance pilot.

SNAP is an important program for millions of families in the United States and cost \$68 billion in fiscal year 2017. State and federal policymakers are interested in developing smart, innovative ways to get a bigger impact for each dollar spent on the SNAP program. Recent research has demonstrated that the once-monthly distribution of SNAP benefits encourages recipients to spend their benefits at the beginning of the month, often all at once. Recipients spend

quickly, clip fewer coupons, and purchase fewer store-brand products (Hastings and Shapiro 2018). This rush to spend may contribute to a monthly cycle of food insecurity where caloric intake and fresh food consumption fall toward the end of the month. Overall nutrition may not increase, and other negative outcomes like crime, school infractions, and hospital admissions may increase as well (Gennetian et al. 2016; Hastings and Shapiro 2018; Hastings and Washington 2010; Seligman et al. 2014; Shapiro 2005; Thompson et al. 1988; Wilde and Ranney 2000). The cost of these outcomes is passed on to state and federal governments through increased spending on Medicaid, policing, incarceration, and use and costs of other safety-net programs.

This behavior by SNAP recipients creates a demand surge for retailers at the beginning of each month. The surge reduces the quality of the retail experience for SNAP clients because they face long lines at check-outs and scarcity of high-demand products. For example, Idaho businesses testified in favor of the transition to staggered SNAP distribution in 2016, citing that it took millions in labor and distribution costs to accommodate the beginning-of-the-month rush (Mohan and Cope 2018). Inevitably, these costs are passed to the consumer, meaning that SNAP recipients may also face higher grocery prices. Once-monthly distribution of SNAP benefits further exacerbates the food insecurity cycle and its associated costs.

RIPL researchers worked in partnership with RI Department of Human Services to use both the RI 360 data lake and grocery scanner data from retail partners to understand how SNAP benefits are spent (Hastings and Shapiro 2018; Hastings, Kessler and Shapiro 2018), and use that knowledge to ascertain if Split-Issuance may be a low-cost policy innovation to support SNAP family nutrition with a product that meets families' needs.

RIPL worked with the RI Department of Human Services and USDA to develop a Split Issuance pilot as a low-cost, high-impact solution to some of the challenges of this monthly cycle. Split Issuance is a distribution change whereby families receive at least two SNAP payments over the course of the month instead of one lump sum; it is a simple and familiar way to diminish the beginning-of-month demand surge and the associated cycle of food insecurity. By providing a steadier stream of income, Split Issuance may support recipients in their effort to smooth their purchases over the course of the month and to purchase more fresh foods.

Developing the pilot required scientific expertise, as well as practical and technical know-how of program operations. By collaborating and developing the RI 360 data lake, we were able to create a fact-based, scientifically designed policy pilot that could be not only provide timely evaluation

of policy impact, but also operate at very low cost using existing benefits distribution technology (RIPL 2017b, 2017c, 2018a).

***Challenge 12. It is crucial to evaluate the impacts of policy innovations so that we continue to learn what works and why, enabling us to test and improve public policy solutions.***

The final stage of fact-based policy is measuring impacts of policy interventions, regulations, and programs. While this is not yet possible for RIPL (because it is still new), three principles should guide impact measurement.

1. *Measuring impact requires long-term commitment to implementing policy change in a measurable way.* Even with adequate data resources and technical expertise, developing a policy prescription may take several months to a year. Cycles for elected officials often last only four years, with first and final years focused on establishing agendas and election campaigns, respectively. Cabinets often turn over every four years even if a sitting governor is reelected. Thus, policy prescriptions are often implemented without a commitment to measuring impact and recalibrating in response to new evidence.

Two approaches can be helpful to encourage future policymakers to evaluate changes implemented by predecessors or to evaluate reallocation of resources as new policy priorities emerge. First, federal agencies can require evaluation of federally funded policies, and can encourage local state and research partnerships to perform the evaluation. Second, foundations giving funds to states to improve policies can likewise require an evaluation of the impact of policy changes.

2. *Impact may be measurable only in the long term, but it is important to optimally identify short-run markers of success.* A policy's progress toward its goals may sometimes be measurable only in the long term. For example, consider the goal of reducing the 36-month recidivism rate. Evaluating a policy to reduce this rate would take at least four years to allow enough incarcerated individuals to be released from prison with sufficient time to measure recidivism rates over a three-year horizon. It may be wise to require this assessment five years post-change, but economists and data scientists can also develop surrogates, or short-run factors that are highly predictive of the long-run outcome of interest. Such methods have been used in technology and business to optimally learn from short-run policy experimentation (Athey et al. 2016). Here again, data resources and collaboration between expert scientists and policymakers can assist in developing the right surrogate models. Federal and grant-making guidelines can provide added incentives to commit states to not only use data

and facts to guide policy changes, but also to measure the impact of those changes toward achieving their goals.

3. *Correct measurement is not one size fits all.* Finally, making a commitment to measuring impact can be difficult. Ideally, we would like to identify causal effects, rather than simply measuring potentially spurious correlations. The gold standard to measure causation is a randomized control trial, which evaluates a policy by randomly assigning it to some individuals to a treatment and not others in order to measure the causal effect of the policy on individual outcomes. However, not all policies can be evaluated with an RCT. In fact, many cannot be evaluated that way for both ethical and mechanical reasons. Instead, expert scientists can consider the available data, collaborate with policymakers to understand what is feasible, and help advise policymakers on how to best evaluate the impact of the policy.

As an example, consider the Rhode2College program described above. Such a program does not naturally lend itself to evaluation based on an RCT for several reasons. Suppose that scholarships were randomly assigned to any student who exceeded a threshold PSAT score. First, it would be difficult to know how scholarship outcomes due to random luck would affect student effort. If scholarship recipients were more likely to go to college than nonrecipients, it would be unclear whether the financial impact of the program or recipients' attitude toward selection or lack of selection was the cause. Second, the RCT does not actually test the effectiveness of a policy that would be implemented at scale. For instance, we might learn from an RCT that Rhode2College did not increase college enrollment because students were notified in private, making it impossible for them to receive peer and teacher affirmation present in a generally available program. For this project, an RCT may not be the correct evaluation tool unless the program can be randomized across districts within a state or the country. Such randomization would require a very large program trial, which may be infeasible. Finally, an RCT may also be ethically and politically infeasible because it would raise vital questions of equity and fairness in administering educational programs.

Fortunately, alternative evaluation strategies such as difference-in-differences, machine learning, and regression discontinuity designs can reliably identify causal impacts in many empirical settings. In addition, surrogate models of college enrollment can be used to glean early insights on likely success of the program and increase the scale adaptively to such scientifically driven priors on likely success as would be done in industry. An approach to fact-based policy that partners data resource construction, expert scientists, and policy practitioners can provide the

support needed to identify and implement the best feasible policy evaluation measures.

## POLICY GOALS SUMMARY

Refining broad policy goals into measurable objectives is a necessary first step to fact-based policy. Here collaboration between policymakers who understand local context, and scientists who can translate that context into data-driven directions for policy innovation is important. Moreover, up-to-date expertise in areas such as machine learning, computer science, and economics is important for distilling data and evidence into robust and reliable directions for improvement.

Data resources like the RI 360 data lake, as well as expertise in modern machine learning and econometric techniques help policymakers translate broad goals into specific, data-driven policies. To help policymakers in time for policy decisions, however, requires the data resource to be quickly accessible and easy to understand. This maximizes the likelihood that research influences and guides actual policymaking, both through researcher partnerships and through governments' own ability to use their data for reporting purposes.

In a little over three years, RIPL and Rhode Island developed and iteratively improved the RI 360 data lake by updating the data lake on a quarterly basis (generating 11 archived research versions), producing 75 policy memos and briefs, and completing over a dozen scientific research papers that helped drive fact-based decisions and innovation in a dozen policies including labor and training, child protection, incarceration, food security, and education.

Note that each project is scalable and applicable to many other states grappling with the same challenges as Rhode Island. By developing standard data resources and data lake structures across localities and states, this knowledge can be replicated and scaled quickly to efficiently and effectively tackle the key societal challenges that most state and local policymakers are striving to address.

Moreover, the data lake can be used by government analysts to support critical internal reporting and cost benefit analysis. For example, the Rhode Island EOHHS used the data lake for reporting on Medicaid enrollment by size of firm employed, and for a report on child protection (EOHHS 2018). This reporting would have required de novo data cleaning, building, matching, and analysis. Drawing on the data lake empowered the agency analytic team to generate needed tables and figures almost immediately.



BOX 4.

## Policy Proposals for Refining Policy Goals and Measuring Progress

- Refine broad policy goals into specific, measurable goals to:
  - Assess where the agency currently is relative to the goal,
  - Develop a path from the current state to the goal that quantifies how progress will be tracked, and
  - Measure success of innovation and policy changed toward reaching the goal.
- Develop specific solutions on the path toward goal achievement.
- Measure the impacts of policy interventions, regulations, and programs with support from:
  - Committing to long-term policy evaluation, supported by the federal government and foundations, which will implement evaluation requirements for programs they fund;
  - Short-run markers of success that aid evaluation when ultimate policy success is only measurable in the long-term; and
  - A wide range of modern econometric techniques that facilitate evaluation even when randomized control trials are infeasible.

---

# Expanding Impact through Fact-Based Policy

---

As a next phase, further developing the model to support fact-based policy efforts more broadly holds the promise of transforming how science and data help guide and innovate efficient and effective public policy that improves lives. Rapid advances in cloud-based secure systems have democratized secure, high-power computing, allowing state and local governments to host data resources like the RI 360 data lake for insights, and to partner with researchers while retaining control of their data and the ability to transparently monitor use.

With curated, reliable, documented data resources optimized for insights by policy-science-engineering core partnerships, expertise and resources can be harnessed across multiple sectors to deliver innovation and insights more efficiently and effectively.<sup>8</sup>

For example, in the domains I am most familiar with (behavioral economics and applied microeconomics), many nonprofit and university-based centers provide scientific expertise to assist policymakers in using data and science to improve policy by connecting networks for academic scientists or engineers with policy needs. Examples of nonprofit groups include groups such as Ideas42, which uses behavioral science evidence to “design scalable ways to improve programs, policies and products in the real world,” and partner with nonprofits, government entities and private business to “make a positive difference to peoples’ lives;” Innovations for Poverty Action, which creates partnerships between academics and policymakers to produce rigorous evidence through randomized controlled trials, and “turn evidence into better programs and policies for the poor;” Code for America, which partners with governments to “redesign public services” and build technology with and for governments in three areas: feeding communities, criminal justice, and economic development.<sup>9</sup>

Nationally-focused university-based groups include MIT-Based J-PAL North America, which “conducts randomized evaluations, builds partnerships for evidence-informed policymaking, and helps partners scale up effective anti-poverty programs;” the Harvard-based Scholars Strategy Network, which “builds communities where research systematically informs policy by connecting local

researchers with key policy players, and trains researchers to communicate research in a way that helps shape policy;” the Harvard Government Performance Lab, which provides “pro bono technical assistance to state and local governments” by sending scholars to collaborate with local policy leaders to “gain insights into the barriers that governments face and the solutions that can overcome these barriers;” and the University of Chicago Urban Labs, which partners researchers with policymakers to “partner with civic and community leaders to identify, test, and help scale the programs and policies with the greatest potential to improve human lives.” The California Policy Lab recently started in the same vein.

Within government, several prominent groups use evidence to support policy through improved cost-benefit analyses, such as the Washington State Institute for Public Policy, which supports cost-benefit analyses for the state legislature using applied policy research and Washington State administrative data; and the Colorado Department of Regulatory Agencies, which conducts cost-benefit analyses of existing and proposed regulations with a focus on occupational licensing (Kleiner 2015). This type of expertise can be an important part of a state government’s capacity to derive the full range of benefits from better data, and their important work can be accomplished at faster speed and with greater impact with access to an integrated, optimized data lake.

All of these models are important approaches for fostering a culture of fact-based policy among their partners. By building integrated, anonymized, and secure data resources for researchers and government partners to access together, states and local governments can provide the resources needed for partners like those above, and many others, to develop and deliver insights at the speed of policy decisions. Moreover, as the need for accountability and transparency grow, data resources can support the large network of nonprofit service organizations, providing vital programs and services to fight poverty and increase opportunity by supporting feedback and evaluation through data resource support to help them measure success and improve their products and services.

---

# Questions and Concerns

---

## **1. Would anonymizing the data prevent agencies and policymakers from accessing data from their own agencies and thereby reduce its usefulness?**

No, agencies and policymakers will still have access to their own original data and will maintain the ability to run models on their own data for program administration. The anonymization process would take place when creating a state-wide database that merges administrative records across agencies. This anonymization is essential for protection of individuals' privacy and does not reduce the usefulness of the data for generating insights that can shape policy.

## **2. Will your proposal, which provides for more and more-detailed outcome measurement, cause those who are implementing policy to shift their focus away from the most important objectives (i.e., lead them to teach to the test)?**

When we cannot directly measure all of the desired outcomes and therefore must use narrower metrics to measure performance, we run a risk of focusing excessively on improving the narrow metric measured. However, one benefit of this model is that the databases proposed here contain large amounts of data from a variety of sources, allowing for measurement of a breadth of outcomes. This attenuates the risk by allowing policymakers to target a broader range of relevant outcomes.

## **3. Would it be possible to merge the state databases with federal data?**

Yes, state databases could be merged with federal data to enhance research capabilities. However, researchers would need to ensure that there is a joint decision-making process among state governments, federal governments, and researchers. All parties would also need to sign the proper legal and privacy agreements.

## **4. Are we at risk of exacerbating structural inequalities when we make more use of data in policy implementation?**

One common and valid concern regarding predictive analytics is that models may be trained on biased or misunderstood data, and thus would serve to amplify rather than ameliorate structural inequalities (Angwin et al. 2016). That said, the use of predictive modeling in the criminal justice system has shown that machine learning can also reduce disparities

between marginalized groups when appropriately and expertly designed. For example, machine learning and predictive modeling approaches can reduce racial disparities in bail decisions, because individual judges may display racial biases that a machine-learning model can adjust for (Kleinberg et al. 2017). More broadly, researchers have begun studying the theoretical aspects of fairness, the impact of predictive modeling on marginalized groups (Hardt, Price, and Srebro 2016; Pope and Sydnor 2011), and other application areas such as credit scoring (Kleinberg, Mullainathan, and Raghavan 2016).

## **5. What sort of reception have policymakers given RIPL?**

The response has generally been very positive. The following quotes were collected from several of our project partners in Rhode Island state and local government.

*Eric J. Beane, former secretary of the Executive Office of Rhode Island Health and Human Services, former deputy chief of staff to the governor of Rhode Island*

“Investing in data-driven and fact-based policy is vital to helping people in need. We need to be innovative and proactive in government to support our communities and families to the best of our ability. Our partnership with RIPL helps us use objective research and data at high speed so we can do just that. RIPL’s approach to data and science for social good could benefit local and state government throughout the country.”

*Scott Jensen, director of the Rhode Island Department of Labor and Training*

“RIPL is a responsive, skilled and trusted partner that shares our commitment to using data to inform policymaking. The invaluable insight and technical assistance provided by RIPL helped the Department fully leverage our data to make better use of federal funds, understand how best to help workers re-enter the workforce, and measure the impact of workforce development programs. We look forward to future collaboration with RIPL as we work to prepare all Rhode Islanders for in-demand jobs.”

*Macky McCleary, administrator of the Rhode Island Division of Public Utilities and Carriers*

“RIPL is making it possible for my agency solve problems with data that we weren’t able to solve before. Their commitment to bringing innovative data solutions to the table means that I’m supported in understanding how policies in my agency and those in other agencies impact or interfere with each other. I am able to make informed policy decisions in ways I wasn’t able to before the RIPL partnership.”

*Trista Piccola, director of the Rhode Island Department of Children, Youth and Families*

“RIPL helps us prove a concept—that we can measure what happens to children DCYF touches and use those facts to transform family policy for the better. RIPL and DCYF proved that not only could we measure these outcomes short- and long-term, but we could measure the impact of our own policies retrospectively and make informed decisions for children and families in need. RIPL researchers have the expertise and ability to aggregate data across multiple systems, which gave us a comprehensive analysis of our state’s children and youth and the effectiveness of particular services. RIPL gave us the support we needed to build a better juvenile justice system for the state and strengthen our home and community-based service array.”

*Arthur Nevins, education policy adviser at the Rhode Island Office of the Governor*

“RIPL has become a trusted partner we can engage to help us design new ways to improve lives for more students meaningfully and measurably. RIPL always get the job done efficiently, effectively, and with a focus on our core needs. They

consistently prioritize how they can deliver the most value for the students and families we serve.”

*Kimberly Paull, director of data and analytics for the Rhode Island Executive Office of Health and Human Services*

“Through partnership, insights, and integrated data, RIPL gave our team the jump start we needed to improve how our state serves whole people, families and communities. RIPL’s ability to quickly gather, clean, connect, and curate data for analysis—and then apply machine learning in collaboration with our teams’ context—gave us a new sense of the possible for data-driven policy-making, academic partnerships, and continuous improvement.

Because of this jump start and the credibility it helped us build through our first major integrated data project, we have now built the capacity to conduct our own integration and analysis in the future, expanding our ability to serve our communities and help families in need.”

*Chief Sid Wordell, executive director of the Rhode Island Police Chiefs’ Association*

“RIPL brings people together around data to solve problems. Working with RIPL over the last couple years has given RI Law Enforcement the opportunity to use RIPL’s integrity and data mining expertise to foster partnerships with a broad range of stakeholders to review often challenging policies. Their professional approach to using data has built trust and promoted transparency without judgement, which has been a refreshing approach for us. I look forward to future partnerships with RIPL as we develop more processes that improve policy with facts.”

---

# Conclusion

---

**R**apid advances in data and computing are driving innovation throughout every aspect of American life, from how we connect with each other, to how we work, make purchases, and manage our homes. The need to harness the power of data and technology to similarly improve public policy impact and efficiency has never been greater.

Fortunately, the potential to improve policy impact and efficiency in meaningful and measurable ways has also

never been greater. The forces driving rapid private-sector technological changes also made it possible to develop public-sector data resources, and to combine those resources with science and practice to improve public policy. Partnerships between public-sector leaders and teams of scientists and engineers are making measurable progress in building the necessary data resources and collaborations to turn data into scientific facts that impact policy and improve lives.

# Appendix

APPENDIX TABLE 1.

## List of Data Sources Combined in the RI 360 Data Lake

<i>Rhode Island Agency</i>	<i>Dataset</i>	<i>Data Start</i>
<b><i>Department of Health</i></b>		
	Birth and vital records, including in-hospital assessments of mother and child health as well as and parent identifiers	1975
	Demographic data (date of birth, gender, ethnicity, address)	1975
	Maternal, Infant and Early Childhood Home Visiting Programs	1997
	Pregnancy Risk Assessment Monitoring System	2002
	KIDSNET (child health registry) and immunization records	1997
<b><i>Executive Office of Health and Human Services</i></b>		
	Medicaid claims	2000
<b><i>Department of Human Services</i></b>		
	Monthly eligibility records for Supplemental Nutrition Assistance Program (SNAP)	1992
	Temporary Assistance for Needy Families (TANF) data	1992
	Comprehensive Community Action Program (CCAP, child care system)	1992
	Case benefit history and income for Department of Human Services programs (SNAP, TANF, Medicaid, CCAP, and general public assistance)	1989
	Child support case-level records	1928
	Electronic benefit transfer cards transaction records (expenditures by card, transaction with vendor information)	2012
	Household expenses of beneficiaries	1990
	General Public Assistance	1992
	Rhode Island State health insurance programs	1992
	Supplemental Security Income and disability benefits records	1974
	All training program information (work training for TANF, SNAP, general public assistance)	1974
<b><i>Department of Education</i></b>		
	Enrollment, demographics, school changes, student address	2004
	Disciplinary infractions	2004
	School staff (teachers), courses, and student information and attendance	2011
	Standardized test scores	2004

AP exam codes and grades	2009
SAT scores	2002
PSAT scores	2009
GED completion, certificate training programs	2004
Charter school lottery data	2012
National Students Clearinghouse post-secondary enrollment and graduation	2004
Public Pre-K applications, class list & attendance	2009

---

***Department of Children, Youth and Families***

Juvenile court, justice, sentence and incarceration data	2000
Child protective services, investigations, placements, foster care, adoption, etc.	2000

---

***Department of Labor and Training***

Wage earnings for individuals employed at RI-based companies	2000
Temporary Disability Insurance claims and payments	1993
Unemployment Insurance claims and payments	1997
Work training data, applications and attendance by person and program	2000
Workers compensation claims	1942

---

***Department of Corrections***

Incarceration information (including variables on families, training programs, in prison behavior)	1995
Probation and parole data	1993

---

***Department of Public Safety***

Driver's license and registration records	Current
Crime History Records	2004
Crime reports, traffic stops, traffic violations, arrests	2004

---

***Department of Environmental Management***

All hazardous waste generators (drug stores, dry cleaners, industry) inspection data	2000
--	------

---

---

# Author

---

## Justine S. Hastings

*Professor of Economics and International and Public Affairs,  
Brown University*

*Founding Director, Research Improving People's Lives*

Justine S. Hastings is a Professor of Economics and International and Public Affairs at Brown University, a Faculty Research Associate with the National Bureau of Economic Research, and the Founding Director of Research Improving People's Lives, a nonprofit research institute using data and science to impact policy and improve lives. Her research expertise spans in Industrial Organization and Public Economics to address important economic and policy questions. She has conducted academic research on topics such as market structure and competition, environment and energy regulation, advertising and consumer protection, consumer financial markets, health care, social safety-net programs, and markets for higher education. Her research employs diverse empirical techniques including field experiments, survey

analysis, machine learning, predictive analytics, analysis of large administrative datasets, and structural demand and supply estimation. Her research was cited in the 2017 Nobel Prize scientific background materials implications, and has been used to shape public policy improvements around the world.

Professor Hastings has served as an expert advisor on the Academic Research Council to the United States Federal Consumer Financial Protection Bureau. She has served as the Managing Editor for the *International Journal on Industrial Organization*, an Editorial Board member of the *Journal of Economic Literature*, and a Co-Editor for the *Journal of Public Economics*. She is a scholar visiting Amazon during the current academic year.

## Acknowledgments

This paper was supported by important contributions from and benefited from conversations with several RIPL team members including Jonelle Ahiligwo, Mintaka Angell, Susan Athey, Anthony Bald, David Berenbaum, Zakary Campbell, Ian Chin, Eric Chyn, Samantha Gold, Mark Howison, Sarah Inman, Victoria Kidd, Ted Lawless, Jeffrey Nelson, Amy O'Hara, Miraj Shah, Jesse Shapiro, Anthony Thomas, John Uclés, Preston White, and Seth Zimmerman. We thank our government partners in the State of Rhode Island for their continued collaboration and commitment to fact-based policy, and Smith Richardson Foundation for financial support.



---

# Endnotes

---

1. I use the term “data lake” to refer to a store of enterprise data that may include raw data but that may also include structured and transformed data optimized for developing insights through reporting, visualization, econometric analysis, forecasting, and machine learning.
2. One example of a useful evidence clearinghouse is the Pew Trusts Results First Clearinghouse Database (Pew-MacArthur Results First Initiative 2015).
3. Public-facing examples include Microsoft Research, JP Morgan Chase Institute, and Zillow Research.
4. RIPL work is funded through foundation and federal government grants for scientific research and applications.
5. In this respect, the needs for fact-based policy in government are similar to those in the private sector. For example, in grocery retail, pricing policy in gasoline may impact demand for in-store products. Firms may therefore want to measure the impact of changes in gasoline prices on grocery revenue, and they may want to combine the customer purchases data with customer demographics to analyze impact across different customer segments.
6. This work was foundation funded and conducted at no direct monetary cost to the state.
7. For example, Atlassian products JIRA and Bitbucket provide integration of task management with code versioning in a code repository; several similar products also exist.
8. RIPL is currently building a template for a standardized RIPL data lake which can be implemented securely in the cloud by any state and locality.
9. In addition, longstanding research institutes such as RAND, Mathematica Policy Research, MDRC, and Research Triangle Institute provide research and technical services to government and for-profit clients world-wide.

---

# References

---

- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016, May 23. "Machine Bias." ProPublica.
- Athey, Susan, and Michael Luca. 2018. "Economists (and Economics) in Tech Companies." Working Paper 25064, National Bureau of Economic Research, Cambridge, MA.
- Athey, Susan, Raj Chetty, Guido W. Imbens, and Hyunseung Kang. 2016. "Estimating Treatment Effects using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index." Working paper, Stanford University, Stanford, CA.
- Autor, David H. 2014. "Skills, Education, and the Rise of Earnings Inequality Among the 'Other 99 percent.'" *Science* 344 (6186): 834–51.
- Avery, Christopher, and Sarah Turner. 2012. "Student Loans: Do College Students Borrow Too Much—or Not Enough?" *Journal of Economic Perspectives* 26 (1): 165–92.
- Bald, Anthony, Eric Chyn, Justine S. Hastings, and Margarita Machelett. 2019. "The Causal Impact of Removing Children from Abusive and Neglectful Homes." Working Paper 25419, National Bureau of Economic Research, Cambridge, MA.
- Bettinger, Eric P., and Bridget Terry Long. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *Quarterly Journal of Economics* 126 (3): 1205–42.
- Beyer, Harald, Justine Hastings, Christopher Neilson, and Seth Zimmerman. 2015. "Connecting Student Loans to Labor Market Outcomes: Policy Lessons from Chile." *American Economic Review* 105 (5): 503–13.
- Billings, John, and Maria C. Raven. 2013. "Dispelling an Urban Legend: Frequent Emergency Department Users Have Substantial Burden of Disease." *Health Affairs* 32 (12): 2099–108.
- Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics* 133 (3): 1163–228.
- Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya Porter. 2018. "Race and Economic Opportunity in the United States An Intergenerational Perspective." Working paper, U.S. Census Bureau, Suitland, MD.
- CNBC. 2015a, June 26. "America's Most Expensive States to Live in 2015." CNBC.
- CNBC. 2015b, June 24. "America's Top States for Business 2015: A Scorecard on State Economic Climate." Last updated July 9, 2018. CNBC.
- Commission on Evidence-Based Policymaking (CEBP). 2017. "The Promise of Evidence-Based Policymaking." Report of the Commission on Evidence-Based Policymaking. Commission on Evidence-Based Policymaking, Washington, DC.
- Conti, Gabriella, James Heckman, and Sergio Urzua. 2010. "The Education-Health Gradient." *American Economic Review* 100 (2): 234–38.
- Deming, David, Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2016. "School Accountability, Postsecondary Attainment, and Earnings." *The Review of Economics and Statistics* 98 (5): 846–62.
- Deming, David, and Susan Dynarski. 2010. "College Aid." In *Targeting Investments in Children: Fighting Poverty When Resources Are Limited*, edited by Phillip B. Levine and David J. Zimmerman, 283–302. Chicago, IL: University of Chicago Press.
- Duarte, Fabian, and Justine Hastings. 2013. "Fettered Consumers and Sophisticated Firms: Evidence from Mexico's Privatized Social Security Market." Working Paper 18582, National Bureau of Economic Research, Cambridge, MA.
- Family Educational Rights and Privacy Act of 1974 (FERPA), Pub. L. 93-380, 88 Stat. 571 (1974).
- Federal Information Security Act of 2002 (FISMA), Pub.L. 107-347, 116 Stat. 2899, 44 U.S.C. ch. 35, subch. III § 3541 et seq.
- Gennetian, Lisa, Roopa Sheshadri, Nathan D. Hess, Aaron N. Winn, and Robert M. Goerge. 2016. "Supplemental Nutrition Assistance Program (SNAP) Benefit Cycles and Student Disciplinary Infractions." *Social Service Review* 90 (3): 403–33.
- Gicheva, Dora, Justine Hastings, and Sofia Villas-Boas. 2010. "Investigating Income Effects in Scanner Data: Do Gasoline Prices Affect Grocery Purchases?" *American Economic Review* 100 (2): 480–84.
- Hanushek, Eric A., and Margaret E. Raymond. 2006. "School Accountability and Student Performance." *Federal Reserve Bank of St. Louis: Regional Economic Development* 2 (1): 51–61.
- Hardt Moritz, Eric Price, and Nathan Srebro. 2016. "Equality of Opportunity in Supervised Learning." Working paper, Cornell University, Ithaca, NY.
- Haskins, Ron. 2011. "Fighting Poverty the American Way." Paper prepared for conference *Anti-Poverty Programs in a Global Perspective: Lessons from Rich and Poor Countries*, Social Science Research Center in Berlin, Germany, June 20–21, 2011. Brookings Institution, Washington, DC.

- Hastings, Justine S., David Deming, Thomas Kane, and Douglas Stalger. 2014. "School Choice and College Attendance: Evidence from Randomized Lotteries." *American Economic Review* 104 (3): 991–1013.
- Hastings, Justine, Ali Hortaçsu, and Chad Syverson. 2017. "Sales Force and Competition in Financial Product Markets: The Case of Mexico's Social Security Privatization." *Econometrica* 85 (6): 1723–61.
- Hastings, Justine S., Mark Howison, Ted Lawless, John Ucles, and Preston White. 2018. "Integrating Administrative Data for Policy Insights." OSF Preprints working paper.
- Hastings, Justine, Ryan Kessler, and Jesse Shapiro. 2018. "The Effect of SNAP on Nutrition." Brown University Working Paper, Providence, RI.
- Hastings, Justine, and Jesse M. Shapiro. 2013. "Fungibility and Consumer Choice: Evidence from Commodity Price Shocks." *Quarterly Journal of Economics* 128 (4): 1449–98.
- . 2018. "How Are SNAP Benefits Spent? Evidence from a Retail Panel." *American Economic Review* 108 (12): 3493–540.
- Hastings, Justine, and Ebonya Washington. 2010. "The First of the Month Effect: Consumer Behavior and Store Responses." *American Economic Journal: Economic Policy* 2 (2): 142–62.
- Hasting, Justine S., and Jeffrey Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics* 123 (4): 1373–414.
- Hastings, Justine S., Seth D. Zimmerman, and Christopher Andrew Neilson. 2015. "The Effects of Earnings Disclosure On College Enrollment Decisions." Working Paper 21300, National Bureau of Economic Research, Cambridge, MA.
- Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. 104-191, 110 Stat. 1936 (1996).
- Heckman, James J, John Eric Humphries, and Gregory Veramendi. 2018. "Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking." *Journal of Political Economy* 126 (S1): S197–S246.
- Hoxby, Caroline, and Sarah Turner. 2013. "Expanding College Opportunities For High-Achieving, Low Income Students." SIEPR Discussion Paper 12-014, Stanford, CA.
- Jensen, Robert. 2010. "The (Perceived) Returns To Education and the Demand For Schooling." *Quarterly Journal of Economics* 125 (2): 515–48.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human Decisions and Machine Predictions." Working Paper 23180, National Bureau of Economic Research, Cambridge, MA.
- Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2016. "Inherent Trade-Offs in the Fair Determination of Risk Scores." Working paper, Cornell University, Ithaca, NY.
- Kleiner, Morris M. 2015. "Reforming Occupational Licensing Policies." Discussion paper 2015-01, The Hamilton Project, Brookings Institution, Washington, DC.
- Lazonick, William, and Mary O'Sullivan. 2000. "Maximizing Shareholder Value: A New Ideology for Corporate Governance." *Economy and Society* 29: 13–35.
- Mann, Cindy. 2013. "Targeting Medicaid Super-Utilizers To Decrease Costs and Improve Quality." CMCS Informational Bulletin, July 24. Department of Health and Human Services, Center for Medicaid and CHIP Services, Baltimore, MD.
- Mann, Cindy. 2014. "Reducing Nonurgent Use of Emergency Departments and Improving Appropriate Care in Appropriate Settings." CMCS Informational Bulletin, January 16. Department of Health and Human Services, Center for Medicaid and CHIP Services, Baltimore, MD.
- Maternal and Child Health Bureau. 2018. "Home Visiting." Accessed December 10, 2018. Human Resources and Services Administration, Rockville, MD. <https://mchb.hrsa.gov/maternal-child-health-initiatives/home-visiting-overview>.
- Mohan, Nishant, and Torrie Cope. 2018, June 18. "Staggering Change: Businesses, Residents Prepare for Changes to Food Stamp Distribution." *Idaho Press*.
- National Center for Education Statistics (NCES). 2013. *The Nation's Report Card: Trends in Academic Progress 2012* (NCES 2013–456). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, Washington, DC.
- . 2017. "Web Tables—Four Years After a Bachelor's Degree: Employment, Enrollment, and Debt Among College Graduates." U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, Washington, DC.
- Norcross, Eileen, and Olivia Gonzalez. 2016. "Ranking the States by Fiscal Condition 2016 Edition." Mercatus Center at George Mason University, Arlington, VA.
- Novy-Marx, Robert, and Joshua D. Rauh. 2009. "The Liabilities and Risks of State-Sponsored Pension Plans." *Journal of Economic Perspectives* 23 (4): 191–210.
- Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25 (1): 159–84.

- Pew-MacArthur Results First Initiative. 2015. “Results First Clearinghouse Database.” Last updated December 19, 2018. Pew Charitable Trusts, Washington, DC.
- . 2017. “How States Engage in Evidence-Based Policymaking: A National Assessment.” Pew Charitable Trusts and MacArthur Foundation, Washington, DC.
- Pope, Devin G., and Justin R. Sydnor. 2011. “Implementing Anti-Discrimination Policies in Statistical Profiling Models.” *American Economic Journal: Economic Policy* 3 (3): 206–31.
- Research Improving People’s Lives (RIPL). 2017a. “Design and Evaluation of a Program To Increase College Enrollment Among Low-Income Students.” Unpublished memo, Rhode Island Department of Education, Providence, RI.
- . 2017b. “How Are SNAP benefits Spent? Evidence from a Retail Panel. Evaluating the Impact of SNAP on Household Spending.” Research Improving People’s Lives, Providence, RI.
- . 2017c. “How Can We Measure Access to Healthy Foods and the Impact Access Has on Low-Income Communities?” Research Improving People’s Lives, Providence, RI.
- . 2017d. “Rhode Island annually funds labor training for more than 1,000 individuals. How can we ensure that these individuals get the most benefit from training?” Research Improving People’s Lives, Providence, RI.
- . 2018a. “The Effect of SNAP on the Composition of Purchased Foods: Evidence and Implications.” Research Improving People’s Lives, Providence, RI.
- . 2018b. “Predictive Template.” Github repository, <https://github.com/ripl-org/predictive-template>. Research Improving People’s Lives, Providence, RI.
- . 2019. “Putting Low-Income Students on the Road to College.” Research Improving People’s Lives, Providence, RI.
- Results for America. 2018. “State Policy.” Results for America, New York, NY. Accessed January 16, 2019. <https://results4america.org/our-work/state-policy/>.
- Rhode Island Department of Children, Youth, and Families (DCYF). 2016. “About Us.” Last modified July 15, 2016. Accessed January 16, 2019. [http://www.dcyf.ri.gov/about\\_us.php](http://www.dcyf.ri.gov/about_us.php).
- Rhode Island Department of Corrections. n.d. “Mission Statement.” Accessed January 16, 2019. <http://www.doc.ri.gov/director/mission.php>.
- Rhode Island Department of Education (RIDE). 2019. “Inside RIDE.” Accessed January 16, 2019. <http://www.ride.ri.gov/InsideRIDE/Overview.aspx>.
- Rhode Island Department of Health. 2019. “About Us.” Accessed January 16, 2019. <http://www.health.ri.gov/about/>.
- Rhode Island Department of Human Services. 2019. “About Us.” Accessed January 16, 2019. <http://www.dhs.ri.gov/AboutUs/index.php>.
- Rhode Island Department of Labor and Training. n.d. “About DLT.” Accessed January 16, 2019. <http://www.dlt.ri.gov/aboutdlt.htm>.
- Rhode Island Department of State. 2019. “About Us.” Accessed January 16, 2019. <http://sos.ri.gov/about-us>.
- Rhode Island Executive Office of Health and Human Services (EOHHS). 2018, December. “Child Maltreatment Prevention Project White Paper.” Unpublished. Executive Office of Health and Human Services, Cranston, RI.
- . 2019. “Our Mission.” Accessed January 16, 2019. <http://www.eohhs.ri.gov/>.
- Rhode Island Police Chiefs’ Association. 2017. “About.” Accessed January 16, 2019. <https://www.ripolicechiefs.org/about-ripca>.
- Rhode Island Public Utilities Commission and Division of Public Utilities and Carriers. 2018. “About the Division.” Last modified May 2, 2018. Accessed January 16, 2019. <http://www.ripuc.org/generalinfo/division.html>.
- RI.Gov. 2018. “Raimondo Unveils Plan To Increase Number of Rhode Islanders With Higher Education.” <https://www.ri.gov/press/view/32564>.
- Ryan, Paul. 2017–18. “H.R.4174 - 115th Congress (2017–2018): Foundations for Evidence-Based Policymaking Act of 2017.” <https://www.congress.gov/bill/115th-congress/house-bill/4174>.
- Seligman Hilary K., Anne F. Bolger, David Guzman, Andrea López, and Kirsten Bibbins-Domingo. 2014. “Exhaustion of Food Budgets at Month’s End and Hospital Admissions for Hypoglycemia.” *Health Affairs* 33 (1): 116–23.
- Shapiro, Jesse M. 2005. “Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle.” *Journal of Public Economics* 89 (2-3): 303–25.
- Thaler, Richard H. 2016. *Misbehaving: The Making of Behavioral Economics*. New York, NY: W. W. Norton.
- Thompson, Frances E., Douglas L. Taren, Elizabeth Andersen, George Casella, Jennifer K. J. Lambert, Cathy C. Campbell, Edward A. Frongillo Jr., and Debra Spicer. 1988. “Within Month Variability in Use of Soup Kitchens in New York State.” *American Journal of Public Health* 78: 1298–301.

- U.S. Department of Justice. 2018. "Criminal Justice Information Services (CJIS) Security Policy." U.S. Department of Justice, Federal Bureau of Investigation, Criminal Justice Information Services Division, Washington, DC.
- Vock, Daniel C. 2015, April. "Gina Raimondo Confronts Rhode Island's Uncertain Future." *Governing Magazine*.
- White House. 2015. "Executive Order: Using Behavioral Science Insights To Better Serve the American People." Office of the Press Secretary. White House, Washington, DC.
- Wilde, Parke E, and Christine K. Ranney. 2000. "The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework." *American Journal of Agricultural Economics* 82 (1): 200–13.



## ADVISORY COUNCIL

GEORGE A. AKERLOF  
University Professor  
Georgetown University

ROGER C. ALTMAN  
Founder & Senior Chairman, Evercore

KAREN L. ANDERSON  
Senior Director of Policy & Communications  
Becker Friedman Institute for  
Research in Economics  
The University of Chicago

ALAN S. BLINDER  
Gordon S. Rentschler Memorial Professor of  
Economics & Public Affairs  
Princeton University  
Nonresident Senior Fellow  
The Brookings Institution

ROBERT CUMBY  
Professor of Economics  
Georgetown University

STEVEN A. DENNING  
Chairman, General Atlantic

JOHN M. DEUTCH  
Institute Professor  
Massachusetts Institute of Technology

CHRISTOPHER EDLEY, JR.  
Co-President & Co-Founder  
The Opportunity Institute

BLAIR W. EFFRON  
Partner, Centerview Partners LLC

DOUGLAS W. ELMENDORF  
Dean & Don K. Price Professor  
of Public Policy  
Harvard Kennedy School

JUDY FEDER  
Professor & Former Dean  
McCourt School of Public Policy  
Georgetown University

ROLAND FRYER  
Henry Lee Professor of Economics  
Harvard University

JASON FURMAN  
Professor of the Practice of  
Economic Policy  
Harvard Kennedy School  
Senior Counselor  
The Hamilton Project

MARK T. GALLOGLY  
Cofounder & Managing Principal  
Centerbridge Partners

TED GAYER  
Executive Vice President  
Joseph A. Pechman Senior Fellow,  
Economic Studies  
The Brookings Institution

TIMOTHY F. GEITHNER  
President, Warburg Pincus

RICHARD GEPHARDT  
President & Chief Executive Officer  
Gephardt Group Government Affairs

JOHN GRAY  
President & Chief Operating Officer  
Blackstone

ROBERT GREENSTEIN  
Founder & President  
Center on Budget and Policy Priorities

MICHAEL GREENSTONE  
Milton Friedman Professor in Economics & the  
College  
Director of the Becker Friedman Institute for  
Research in Economics  
Director of the Energy Policy Institute  
University of Chicago

GLENN H. HUTCHINS  
Co-founder, North Island  
Co-founder, Silver Lake

JAMES A. JOHNSON  
Chairman, Johnson Capital Partners

LAWRENCE F. KATZ  
Elisabeth Allison Professor of Economics  
Harvard University

MELISSA S. KEARNEY  
Professor of Economics  
University of Maryland  
Nonresident Senior Fellow  
The Brookings Institution

LILI LYNTON  
Founding Partner  
Boulud Restaurant Group

HOWARD S. MARKS  
Co-Chairman  
Oaktree Capital Management, L.P.

MARK MCKINNON  
Former Advisor to George W. Bush  
Co-Founder, No Labels

ERIC MINDICH  
Chief Executive Officer & Founder  
Eton Park Capital Management

ALEX NAVAB  
Former Head of Americas Private Equity  
KKR  
Founder, Navab Holdings

SUZANNE NORA JOHNSON  
Former Vice Chairman  
Goldman Sachs Group, Inc.

PETER ORSZAG  
Head of North American M&A &  
Vice Chairman of Investment Banking  
Lazard Freres & Co LLC

RICHARD PERRY  
Managing Partner & Chief Executive Officer  
Perry Capital

PENNY PRITZKER  
Chairman & Founder, PSP Partners

MEEGHAN PRUNTY  
Managing Director, Blue Meridian Partners  
Edna McConnell Clark Foundation

ROBERT D. REISCHAUER  
Distinguished Institute Fellow &  
President Emeritus  
Urban Institute

ALICE M. RIVLIN  
Senior Fellow, Economic Studies  
Center for Health Policy  
The Brookings Institution

DAVID M. RUBENSTEIN  
Co-Founder & Co-Executive Chairman  
The Carlyle Group

ROBERT E. RUBIN  
Former U.S. Treasury Secretary  
Co-Chair Emeritus  
Council on Foreign Relations

LESLIE B. SAMUELS  
Senior Counsel  
Cleary Gottlieb Steen & Hamilton LLP

SHERYL SANDBERG  
Chief Operating Officer, Facebook

DIANE WHITMORE SCHANZENBACH  
Margaret Walker Alexander Professor  
Director  
The Institute for Policy Research  
Northwestern University  
Nonresident Senior Fellow  
The Brookings Institution

STEPHEN SCHERR  
Chief Executive Officer  
Goldman Sachs Bank USA

RALPH L. SCHLOSSTEIN  
President & Chief Executive Officer, Evercore

ERIC SCHMIDT  
Technical Advisor, Alphabet Inc.

ERIC SCHWARTZ  
Chairman & CEO, 76 West Holdings

THOMAS F. STEYER  
Business Leader & Philanthropist

LAWRENCE H. SUMMERS  
Charles W. Eliot University Professor  
Harvard University

LAURA D'ANDREA TYSON  
Professor of Business Administration &  
Economics  
Director  
Institute for Business & Social Impact  
Berkeley-Haas School of Business

---

JAY SHAMBAUGH  
Director

## Highlights

Policymakers at the state and local levels tackle some of the toughest problems facing society. To make measurable progress in solving these problems, public policy needs to be effective, efficient, and evidence based. Justine Hastings of Brown University draws on her experiences founding Research Improving People's Lives (RIPL), a nonprofit research-policy partnership in Rhode Island that aims to make state policy more fact based and more effective through the creation of an integrated database of state administrative data and derived tables. She provides best practices for creating the database and using it effectively to derive policy insights.

## The Proposal

**Create an integrated database of all state administrative data.** State governors will organize computer scientists, economists, engineers, and policy experts who will in turn collect, document, encrypt, anonymize, and merge data from state agencies' administrative databases to optimize the usefulness of the resulting database. The database will conform to best practices for data privacy, which are often more stringent than state and federal privacy laws.

**Optimize the database for policy insights by creating a data lake.** Database engineers will create a data lake by transforming all back-end, raw data into useable data for measuring policy impact by creating a relational database with derived tables.

**Set appropriately refined policy goals, develop concrete policy solutions, and continually measure short- and long-term impacts of policy changes and progress toward goals.** Policymakers and research experts will collaborate to refine broad policy goals into specific, measurable goals. This requires that they assess where the agency currently is relative to the goal, develop a path from the current state to the desired outcome, including quantifying how progress will be tracked, and measure success and impacts of innovation and policy changes toward reaching the goal.

## Benefits

By developing public sector data resources and combining those resources with the appropriate scientific and practical expertise, Hastings shows that states can substantially improve public policy. Data resources like those proposed here help policymakers translate broad goals into effective, data-driven policies, which in turn deliver strong value to state governments and taxpayers. Partnerships between public sector leaders and teams of scientists and policy experts are already beginning to build the necessary data resources and collaborations to make better policy.



1775 Massachusetts Ave., NW  
Washington, DC 20036

(202) 797-6484

BROOKINGS



Printed on recycled paper.

[WWW.HAMILTONPROJECT.ORG](http://WWW.HAMILTONPROJECT.ORG)