# THE HAMILTON PROJECT

Advancing Opportunity, Prosperity, and Growth

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

JANUARY 2019

## 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.

### Hamilton Project Updates

www.hamiltonproject.org

www.facebook.com/hamiltonproject

www.twitter.com/hamiltonproj

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

*Whether they are attempting to alleviate poverty, increase economic opportunity, or improve education and health care,* state and local policymakers work to 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.

In a new Hamilton Project proposal, Justine Hastings of Brown University and Research Improving People's Lives (RIPL) draws on her experience founding RIPL, a nonprofit research–policy partnership in Rhode Island that aims to make state policy more fact based and effective. Through the author's collaboration with the state of Rhode Island, she has identified several challenges that policymakers face in successfully implementing fact-based policies: developing effective and secure data resources for insights, collecting the necessary technological resources and expertise, and reliably defining and measuring program success. Hastings's proposal offers solutions to these challenges based on her experience and outlines best practices for the development of policy-relevant data resources.

## The Challenge

Fact-based policy is essential to making government more effective and more efficient. This is particularly true, the author contends, at the state and local levels where policymakers are increasingly looking for innovations that deliver results at lower costs and where resources for measuring benefits and costs have not always been available.

In the private sector, as Hastings notes, advances in technology have allowed firms to make greater use of comprehensive data and state-of-the-art technology. These advances have yielded insights about what people need, allowed firms to develop and innovate to meet those needs, and facilitated measurement of key outcomes. Firms are increasingly building data resources and employing economists, statisticians, and data scientists to measure and improve their products and services.

Hastings proposes state governments use a similar strategy to produce high-impact results for the communities they serve. Steeply falling prices of computing and other technological advances have made possible new ways to measure impact and deliver products and services that meet a wide range of needs. See box 1 for examples of governments supporting fact-based policy.

**BOX 1.**

### Recent Efforts to Support Fact-Based Policy

Government is beginning to respond to the new technological possibilities. During his second term in office, President Obama issued executive orders to incorporate data and behavioral economics in policy design and evaluation. Senator Patty Murray and then-Speaker Paul Ryan established a bipartisan commission to investigate ways to promote data-driven policy. One of the recommendations the commission presented in its September 2017 final report was the creation of data infrastructure to facilitate the development of more-effective public policy. Governor Gina Raimondo of Rhode Island has supported the development of data resources, and has put those resources to work in partnership with scientific research teams led by the author.

Many states could benefit from more extensive use of data and evidence when making policy. The Pew-MacArthur Results First Initiative has reviewed state governments' use of cost-benefit analysis across various policy areas, and has found wide variation in state practices (see figure 1).

Hastings explains that the first necessary condition for making fact-based policy is a data resource that can generate useful, policy-relevant facts. Though data are ubiquitous within and across government agencies, they usually exist in siloed agencies and, because they have not been collected with program evaluation in mind, are not in a useable format for generating policy insights. Data resources need to be built to support fact-based policy, but in a way that complies with federal and state privacy laws as well as best practices for generating robust and reliable insights and protecting data privacy—which may go beyond simple legal compliance.

Creating such a resource requires collaboration between practitioners and scientists. The author notes that within private industry such research–practitioner collaborations often happen in house: firms hire computer scientists and economists to work with business leaders. While such hiring may not be feasible for government due to budgetary limitations and other restrictions, partnerships with external researchers are often possible. Crucially, practitioners know the ins and the outs of data fields and how they should be interpreted. Collaboration with researchers is needed in turn to determine what data formats are necessary for research insights, how to translate policy goals into data-driven insights, and how to construct the technical methodology to transform back-end data into an integrated system, all while complying with federal and state privacy laws.

FIGURE 1.
## Cost-Benefit Analysis Use by State



Number of policy areas using cost-benefit analysis
0   1   2   3   4
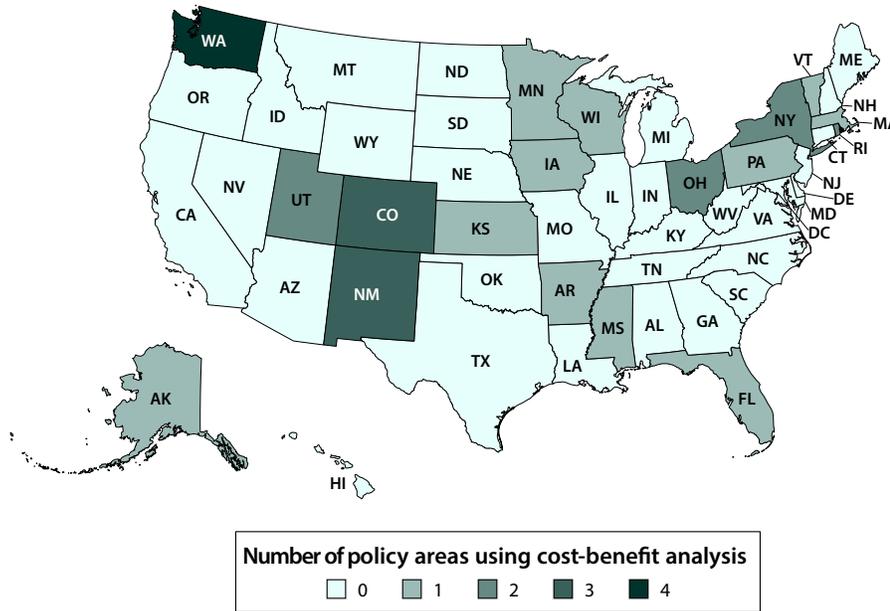
Source: Pew-MacArthur Results First Initiative 2017; author's calculations.

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 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.

THE HAMILTON PROJECT
BROOKINGS

# A New Approach

Based on the challenges outlined above, Hastings proposes that states build comprehensive, integrated databases, transform those databases into data lakes—databases that, in addition to raw data, also contain derived tables that are optimized for developing insights—and use the insights from this work to sharpen policy goals, create policy solutions, and measure progress against those goals. The comprehensive scope of this work could make it possible for researchers and policymakers to work together to address more effectively a wide range of challenges, such as lowering health-care costs, reducing the impact of the opioid epidemic, improving worker training programs, connecting dislocated workers to benefits, helping families become more

food secure, optimizing energy policy for low-income families, helping children reach proficiency on reading and math tests, and closing the college achievement gap. Modern, secure, cloud-based computing allows government to retain control of its data with the highest security standards while harnessing cutting-edge computing resources to partner efficiently and confidently with external scientific teams.

## Building an Integrated Database

Hastings's research team and Rhode Island partners worked to build a prototype model: an anonymized, integrated database that draws from the administrative records of state agencies across Rhode Island. Drawing on her experience with that work, the author proposes that other states adhere to the following best practices as they construct their own integrated databases:

- **Computer scientists, economists, engineers, and policy experts should work together to reorganize and prepare data** from administrative datasets to optimize the database for policy insights. Throughout state government, data are typically recorded for program administration purposes and not for research insights. Program administrators and computer scientists need to keep research interests in mind as they create and curate key data sets that contain information that is useful for a variety of purposes.

- **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

## Learn More about This Proposal

This policy brief is based on the Hamilton Project policy paper, "Fact-Based Policy: How Do States and Local Government Accomplish It?," which was authored by

JUSTINE HASTINGS

Brown University
Research Improving People's Lives

can reliably use variables to draw the appropriate policy conclusions from their analyses.

- **Data need to be anonymized, housed, and used in ways that protect confidentiality** and meet (and exceed) the security standards of federal and state privacy laws governing the data. All data should be anonymized, and personally identifiable information should be removed from the data, retaining only the minimum amount of information necessary to accomplish the research goals. Once the data are secured, only approved and monitored access should be permitted. These steps should be implemented even if federal laws do not require them.

- **Disconnected datasets in agency silos should be encrypted and merged** using anonymized identifiers that meet privacy requirements and safeguard the security of individuals' records. Once the anonymized identifiers are created, encrypted personal identifiers should be removed from the database.

- **In order to comply with federal privacy laws, data should be made available only for approved projects** that further the mission of the relevant state agency. Projects need to be defined and tied to specific data tables, and access restrictions must be implemented so that only the data needed for each approved project is accessible by project researchers and analysts. This can be facilitated by modern, secure, cloud-based computing.

## Creating a Data Lake: Developing the Data Resource for Policy Insights

Once the components of the database have been properly collected, researchers and computer scientists need to make the data as useful as possible to researchers. Hastings suggests that this can be done by creating a data lake like RI 360, that she and her team created in Rhode Island, which supports ongoing research that aims to alleviate poverty and increase economic opportunity. This data resource consists of derived tables of the most-used data sources in addition to the raw administrative data. See box 2 for more details.

To create a data lake, database administrators must do the following:

- **Transform all back-end, raw data into useable data for measuring policy** impact by creating a relational database with expertly constructed derived tables. Derived tables would consist of the data fields most commonly needed across projects and over time. As new needs are identified, new permanent derived tables can be added to the data lake.

- **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.

### Roadmap

- Policy leaders 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. This will be done in a manner that complies with best practices for data security and privacy, which are often more stringent than requirements in state and federal privacy laws.

- Once the database has been created, database engineers will create a data lake by transforming all back-end, raw data into useable data for measuring policy impact, creating a relational database with derived tables. Because many agencies have similarly formatted data across states, scripts produced in one state can be easily adapted to serve another.

- 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 of efforts toward reaching the goal.

- Using the data lake, economists and policymakers will assess the impacts of policy interventions, regulations, and programs.
  ◦ The federal government and foundations will support measurement efforts by implementing evaluation requirements for programs they fund.
  ◦ Economists and policy experts will identify short-run markers of success that facilitate evaluation when relevant outcomes are observable only in the long term.

Administrators will need to have a process for automatically incorporating new data in a way that is carefully documented.

- **Implement standards for research and insights** in order to enhance robustness, reproducibility, and scalability. Teams may use data to produce insights that other researchers can learn from and use. Standardization of documentation, code, and processes across projects can help ensure that the knowledge developed is scalable.

- **Develop a formal chartering process with government partners** to define the project goals, the data fields needed, the key points of contact in the agencies and scientific teams, and the deliverables and timeline. Projects involve multiple parties—researchers, policymakers, and agencies—requiring careful coordination.

## The RI 360 Data Lake Case Study

The RI 360 data lake contains information on conditional cash transfer programs such as Temporary Assistance for Needy Families, the Supplemental Nutrition Assistance Program, and the Child Care Assistance Program; vital records data on births, deaths, and marriages; medical data such as immunization records and Medicaid claims; labor data on employment and earnings, unemployment insurance, disability insurance, work training programs, and workers' compensation; and information about student enrollment and test scores. Administrative records often go back for decades, permitting analysis of past program changes.

Records are at the individual level, but are anonymized so that individuals cannot be identified. RIPL researchers designed a data importation process that automatically encrypts sensitive information, allowing records to be joined across agencies based on only an anonymized identifier. Security and automatic auditing and logging protocols further prevent unauthorized access to the RI 360 data lake or downloads of data from the system. Users can remove only research results for approved projects. Finally, code was developed to produce tables for high-level insights, transforming all back-end, raw data into useable data for measuring policy impact.

This database structure can be built in any state since all states have similar underlying data structures. For example, unemployment insurance claims records are stored in similar formats to fulfill reporting requirements. Similarly, all states must keep corrections and incarceration data. In many cases, vendors keep the same database format across states (e.g., many Electronic Benefit Transactions in the United States are managed by the same vendor). The significant up-front cost that RIPL incurred in building an integrated database in Rhode Island can thus facilitate the quick and accurate scaling of data systems necessary to build equivalent databases in other states.

### Setting Policy Goals, Developing Policy Solutions, and Measuring Progress

Policymakers need to define their policy goals and refine them into measurable pathways. This requires that policymakers define specifically what it means to achieve the goal or answer the question. Refining a broad goal into a specific, measurable goal often requires collaboration between policymakers and researchers to define measurable outcomes that reflect

policymaker intent, and that incorporate local context, constraints, and needs. Researchers can also help bring to light specific policy solutions based on research that examines the causal impact of policy on outcomes. Finally, researchers and data scientists can help set up useable measurement systems to track progress against the goal. According to the author, to best take advantage of the data resources, policymakers should do the following:

- **Refine broad policy goals into specific, measurable goals** by

  ◦ assessing where the agency currently is relative to the goal;

  ◦ developing a path from the current situation to the desired outcome, including quantifying how progress will be tracked; and

  ◦ measuring success of innovation and policy changes toward reaching the goal.

- **Develop specific solutions** on the path toward goal achievement. In many cases, researchers can employ rigorous evidence that helps direct policy solutions to better achieve the policymaker's goals.

- **Measure the impacts of policy interventions, regulations, and programs** by

  ◦ committing to long-term policy evaluation, supported by the federal government and foundations, that will implement evaluation requirements for programs they fund;

  ◦ identifying short-run markers of success even when policy success is measurable only in the long term; and

  ◦ using tests and measurement techniques that are appropriate for the program context.

# Conclusion

Rapid advances in data and computing are driving innovation throughout American life, from how we connect with each other, to how we work, make purchases, and manage our homes. A similar potential exists for harnessing the power of data and technology to improve public policy impact and efficiency.

Hastings shows that by developing public sector data resources and combining those resources with the appropriate scientific and practical expertise, states can substantially improve public policy, reliably and confidently. Partnerships between public sector leaders and teams of academic scientists are already beginning to build the necessary data resources and collaborations to inform policy and improve lives.

# Questions and Concerns

## 1. 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 the desired outcomes and therefore must use narrower metrics to measure performance, there is a risk of focusing excessively on improving the narrow metric measured. However, one of the biggest benefits of this model is that the proposed databases contain large amounts of data from a variety of sources, allowing for measurement of a breadth of outcomes. This breadth of information attenuates the risk by allowing policymakers to target a wide range of relevant outcomes.

## 2. 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.

# 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.