





UNLOCKING RESPONSIBLE ACCESS TO DATA to Increase Equity and Economic Mobility

Markle Foundation, March 2021 With support from the Bill & Melinda Gates Foundation



ACKNOWLEDGMENTS

This report was compiled by the Markle Foundation with generous support of the Economic Mobility and Opportunity team from the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the foundation. With appreciation for the support and thought partnership from Ryan Rippel, Michael George, and Kosar Jahani of the Economic Mobility team.

We are deeply grateful to the many experts who contributed to our research, with special thanks to David Marsh, Denis McDonough, and Lynn Overmann for their leadership of this project, and to Nick Hart, Nancy Potok, and Stefaan Verhulst for their engagement and support.

Our thanks go out to Desmond Amuh, Byron Auguste, Angela Bowens-Gamble, Steven Brown, Jonathan Burks, Sylvia Burwell, Scott Cheney, Courtney Coffin, Carrie Conaway, Dennis Culhane, Michael Deich, Frank DiGiammarino, Diana Farrell, Heidi Goldberg, Robert Groves, Darrick Hamilton, Amy Hawn Nelson, Brian Harris-Kojetkin, Nathaniel Hendren, Juliana Herman, Lesley Hirsch, Bill Hoagland, Anna Hui, Elisabeth Jacobs, Tom Kalil, Tyler Kleykamp, Paige Kowalski, Julia Lane, Rafael López, Bruce Madson, Chuck Marr, David Marsh, Ted McCann, Denis McDonough, Shantel Meek, Alejandra Montoya-Boyer, Cecilia Muñoz, Amy O'Hara, Lynn Overmann, Jennifer Pahlka, David Park, George Putnam, Arloc Sherman, Paige Shevlin, Julia Stoyanovich, John Thompson, and Nicole Wong for their expert insights and contributions.

TABLI	E OF CONTENTS	
ACKN	OWLEDGMENTS	. 2
Executi	ive Summary: The Need to Unlock Data to Inform Equity and Economic Mobility	. 5
Met	hodology	. 5
Key	Takeaways	. 5
PART (RESEA	ONE: UNDERSTANDING AND ADDRESSING END USER NEEDS: POLICYMAKERS, ARCHERS & THE PUBLIC	. 7
Polie	cymaker User Needs	. 7
0	Increased Data Equity is Critical to Addressing Racial and Ethnic Disparities	. 7
0	Data on program effectiveness to enable comparisons and inform tradeoffs	. 8
0	The Unique Needs of State, Local and Tribal Policymakers	. 8
Rese	earcher User Needs	. 9
0	Data Discoverability, Access & Descriptors	. 9
0	Timely Data Analysis to Inform Policy	10
0	Public User Needs: Increased Education and Engagement	10
PART	TWO: CHALLENGES AND OPPORTUNITIES IN THE CURRENT DATA LANDSCAPE	11
Imp	roving the Quality and Accessibility of Federal Data	11
Key	Takeaways	11
0	Federal reforms will require long-term effort	11
0	The federal government needs increased data-linkage and sharing capacity	12
0	The U.S. Census Bureau should reduce barriers to data access	12
0	Rigorous privacy protections are compatible with increased data sharing, but trust in federal processes has eroded.	13
0	Effective communication around the risk of data misuse should also emphasize the benefits of privacy-protected data sharing	14
0	Federal and state agencies need consistent, predictable funding to improve data collection and sharing.	15
0	Strategies to improve the accessibility and quality of federal data	16
0	Proposed Strategy: Increase Expert Capacity in Federal Government to Advance Federal Data Reforms:	16
0	Proposed Strategy: Reduce barriers to linked, longitudinal federal data	17
0	Proposed Strategy: Create strong external demand signals.	17
0	Proposed Strategy: Expand research leveraging federal data and democratize access:	18
0	Proposed Strategy: Continue progress on sharing open, machine-readable data.	18



INCRE	ASING THE ACCESSIBILITY AND USE OF STATE, LOCAL, AND TRIBAL DATA	19
Key	Takeaways	20
0	State government data collaborations offer great promise	20
0	State administrative data shared with the federal government should be enhanced with federal da & returned to improve state, local, and tribal insights	ıta 20
0	Data on Benefits Delivery are Critical in Response to Economic Shocks	20
0	State reporting requirements should move beyond a primary focus on compliance	21
Stra	tegic Opportunities at the State, Local, and Tribal Level	21
0	Proposed Strategy: Increase capacity and support expansion of regional data collaboratives	21
0	Proposed Strategy: Increase state government data capacity	22
0	Proposed Strategy: Increase research and analysis focused on improving benefits delivery and evaluating program impact	22
INCRE AND I	EASING EQUITY IN GOVERNMENT DATA IS CRITICAL TO ADDRESSING DISPARITY NEQUITIES	23
Key	Takeaways	23
0	Increasing data disaggregation and mitigating bias in data are critical steps to increasing data equity	23
0	Statistical methods for data cleaning may introduce bias	23
0	Addressing data equity requires community-centered engagement	24
Stra	tegic Investment Opportunities to Increase Data Equity	25
0	Proposed Strategy: Support for research on the most effective methods for public engagement ar education:	nd 25
0	Proposed Strategy: Support research to reduce disparities and increase equity:	25
0	Proposed Strategy: Increase transparency around data use and oversight of data-based decision- making tools	26
LEVE	RAGING NEW DATA AND CREATING NEW ECONOMIC MEASURES	26
Key	Takeaways	26
0	Broader measures of economic well-being could help advance mobility	26
0	Non-traditional data offer new opportunities to provide more timely and granular insights	27
Stra	tegic Opportunities to Leverage Non-Traditional Data	28
0	Proposed Strategy: External researchers and stakeholders should demonstrate new or high value uses of data to explore new economic measures, create new data "public goods", and inform federal processes	28
PART	THREE: A STRATEGIC ROADMAP: ACCELERATING THE USE OF DATA TO ADVANCE	20
EQUIT	Y AND MOBILITY	29



Executive Summary: The Need to Unlock Data to Inform Equity and Economic Mobility

Economic mobility remains elusive for far too many Americans and has been declining for several decades. A person born in 1980 is 50% less likely to earn more than their parents than a person born in 1950 is. While all children who grow up in low-opportunity neighborhoods face mobility challenges, racial, ethnic, and gender disparities add even more complexity. In 99% of neighborhoods in America, Black boys earn less, and are more likely to fall into poverty, than white boys, even when they grow up on the same block, attend the same schools, and have the same family income. In 2016, a Pew Research study found that the median wealth of white households was ten times the median wealth of Black households and eight times that of Hispanic households. The COVID-19 pandemic has further exacerbated existing disparities, as communities of color suffer higher exposure and death rates, along with greater job loss and increased food and housing insecurity.

Reversing this overall decline to address the persistent racial, ethnic, and gender gaps in economic mobility is one of the great challenges of our time. Some progress has been made in identifying the causes and potential solutions to declining mobility, yet policymakers, researchers, and the public still lack access to critical data necessary to understand which policies, programs, interventions, and investments are most effective at creating opportunity for students and workers, particularly those struggling with intergenerational poverty. Data collected across all levels of governments, nonprofit organizations, and private sector companies can help answer foundational policy and research questions on what drives economic mobility. There are promising efforts underway to improve government data infrastructure and processes at both the federal and state levels, but critical data often remains siloed, and legitimate concerns about privacy and civil liberties can make data difficult to share. Often, data on vulnerable populations most in need of services is of poor quality or is not collected at all.

To tackle this challenge, the Bill and Melinda Gates Foundation (BMGF) and the Markle Foundation (Markle) spent much of 2020 working with a diverse range of experts to identify strategic opportunities to accelerate progress towards unlocking data to improve policymaking, answer foundational research questions, and ensure that individuals can easily and responsibly access the information they need to make informed decisions in a rapidly changing environment.

Methodology

Over the course of the Spring and Fall of 2020, the Markle team conducted five expert working group sessions and held multiple one-on-one calls with experts to surface challenges and opportunities in improving the government data ecosystem. The expert working groups included a bipartisan group of policymakers from federal, state, and local governments, with experience in the legislative and executive branches, including running agencies and administering benefits, programs, and services. Our team also engaged with a range of researchers, data analysts, and nonprofit organizations working to understand pathways to mobility and to evaluate the impact and effectiveness of policies, especially across low-income and marginalized populations.

Key Takeaways

The experts agreed that federal data is difficult to discover and access, is not timely or geographically granular enough, and is not sufficiently disaggregated by race, ethnicity, gender, and other key demographic variables to measure and address disparity. Policymakers highlighted their struggle to understand which populations in which locations are most in need of support; what kinds of support, programs, or policies will be most effective at increasing mobility; and whether implemented policies have the intended impact. Working to answer core questions about economic mobility, researchers raised challenges in understanding what data are most relevant, what gaps or biases may exist in the data, and



how to navigate long and complex processes to access federal data. Too often, the needs of the public and communities are not prioritized in government decisions, such as which data to collect and disseminate, how to protect privacy and control access, and which research questions to prioritize. This lack of engagement exacerbates the general public's lack of education regarding data use and how it may benefit them, increasing public distrust and limiting opportunities to leverage data to improve outcomes for individuals and communities. Ensuring rigorous privacy protections while also ensuring equitable access to data was a central theme throughout the expert engagements.

The experts identified five major areas of effort where reducing barriers and increasing capacity will significantly advance privacy-protected data sharing, help address disparity and inequality, and improve knowledge on how to increase economic mobility:

- Improving the quality and accessibility of federal data, particularly to facilitate targeted policymaking, expand rigorous research, and make the government more responsive to external demand signals;
- Increasing equity in federal data to better understand and address racial disparity and other inequities;
- Improving the accessibility and use of state data to obtain better insights into programs and benefits provided at the state level and to allow state and local policymakers to better understand and meet needs across geographies and populations;
- Increasing engagement with the public and community stakeholders on data collection, use, and reuse; and
- Leveraging new data and creating new economic measures.

In this paper, we summarize more detailed takeaways across these focus areas, identify the key challenges and opportunities in each, and propose a range of strategies that could accelerate a more equitable, comprehensive, and accessible system of privacy-protected data sharing to improve mobility and opportunity.

In Part One we share the needs of key end users of government data: policymakers at all levels of government, researchers, and the public.

In Part Two of the paper, we share a broader landscape of the current data ecosystem, share key takeaways on challenges and opportunities that emerged from the expert working groups, and share detailed recommendations for improvements.

In Part Three we share an eight-point strategic roadmap to increase data capacity at the federal, state, and local levels; to increase data equity; to improve public and community data; and to create external demand signals.



PART ONE: UNDERSTANDING AND ADDRESSING END USER NEEDS: POLICYMAKERS, RESEARCHERS & THE PUBLIC

Timely, comprehensive information is critical to improving upward mobility, both in times of economic growth and during economic shocks. Data can provide insights into which existing policies and programs are working, where gaps in knowledge or policy can be filled by research and innovation, and how to convey information to the public to empower effective decision-making. In order to develop meaningful strategies for data system improvements, we started by exploring the needs of three key end-users of government data: policymakers, researchers, and the public. We note that an improved government data ecosystem that more broadly shares more timely, granular, and equitable privacy-protected data will offer benefits far beyond these three end-user groups, helping to guide nonprofit and community organizations working to advance mobility, spurring business investment and innovation, and continuing the democratization of data to a wide and diverse audience.

Policymaker User Needs

We embarked on this effort at the beginning of the COVID-19 crisis and focused our initial conversations with policymakers on the need for data to understand and address economic crises. Policymakers at all levels of government want data to guide their decisions, to measure short-term impact and longer-term outcomes, and to help them shift resources in response to changing circumstances on the ground. Policymakers are also increasingly aware of the structural barriers created by systemic racism, the chronic underinvestment in marginalized communities, and major disparities in wealth and opportunity—all of which have been starkly revealed and compounded by the COVID-19 pandemic.

Timeliness and granularity of data to support swift and responsive policymaking was a key concern, but policymakers also highlighted the relative lack of data or research on the impact of policies or programs over time—data which would directly inform future policymaking. As an example of this challenge, the policymakers pointed to the major investments in green jobs and infrastructure made in the American Recovery and Reinvestment Act passed in response to the economic crisis of 2009 and the inability to assess the cumulative effects of those investments on the environment or on the job market. In this paper, we address several key themes highlighted by the policymakers and experts with whom we spoke, including:

- The need for increased data equity to inform policies and measurably reduce disparities;
- The need for data on program and policy effectiveness to make informed comparisons across policies; and
- The different challenges and opportunities in leveraging data faced by federal, state, tribal, and local governments, including capacity and data infrastructure.

o Increased Data Equity is Critical to Addressing Racial and Ethnic Disparities

Increasing data disaggregation by key demographics and geographies is critical to helping policymakers understand and address existing disparities. Federal data and statistics are often not disaggregated by race, ethnicity, gender, disability, income, or other key demographic variables, presenting significant challenges to measuring and meeting needs across diverse populations. These challenges are exacerbated in times of economic shock or other major disruptions when data from past years may not represent changing realities. The lack of demographic and geographic granularity impacts all phases of the policymaking process, from targeting policy responses, to understanding whether benefits are reaching intended populations and whether policies are achieving their intended impact.

For example, it was apparent early in the COVID-19 crisis that communities of color and low-income families were suffering disproportionately, both economically and from increased infection and death



rates. Policymakers looking to provide swift support would have benefitted from knowing, in near realtime, how many people had lost their jobs, how many had applied for unemployment benefits, whether they had received their benefits, and what other economic repercussions, like food or housing insecurity, were occurring. Instead, unemployment insurance data lagged by several weeks and was aggregated to the county or state level. Even as emergency legislation increased and supplemented or provided new benefits, state unemployment systems could not quickly convert the new policy into payments, and many people struggled to get through the application processes, delaying critical aid.

• Data on program effectiveness to enable comparisons and inform tradeoffs.

Policymakers also identified the need to make "apples-to-apples" comparisons across several dimensions—geography, populations, and different programs—as they develop budgets, write legislation, and deliver services. This is particularly critical for state and local policymakers, who often must operate under balanced budget requirements and need data to compare relative costs and outcomes of different policies to help make informed tradeoffs. Even without the formal constraints of balanced budget requirements, budget tradeoffs are still a reality in negotiations at the federal level, where challenges are further compounded by the vast range of activities the federal government funds. One federal budget expert gave the example of having to decide between increased funds for night-vision goggles for the military or increased funds for a promising workforce training problem, without knowing the benefit of the marginal dollar of funding one over the other.

o The Unique Needs of State, Local and Tribal Policymakers

States play a unique role in the data ecosystem, often serving as an intermediary between federal and local governments, overseeing the delivery of many federal safety-net benefits programs, workforce funds, and training funding, and collecting a wide range of administrative data. While implementing these programs, state, local, and tribal governments are also responsible for procuring and maintaining the core data and technology infrastructure to administer the programs. Local and tribal governments play a valuable role as the formal or informal "last mile" of service delivery and are often more attuned to varying and diverse community needs and have valuable insights on how to improve delivery to diverse groups. Despite this central role in oversight, administration, and delivery of key programs, state, local, and tribal governments often struggle with out-of-date technology and data infrastructure that make implementing policy changes or sharing data difficult.

Many state, local, and tribal governments also lack sufficient internal capacity or expertise to conduct data analysis to inform policy or evaluate their own programs. In many state agencies, the only dedicated full-time data staff are focused on complying with federal data reporting requirements. Data capacity varies even more across local and tribal governments, which face numerous challenges in hiring personnel with expertise in data, governance, and technology procurement. Additionally, many rural and tribal governments continue to grapple with inconsistent or non-existent Internet access for their constituents, further limiting access to key programs and benefits.

Despite these capacity constraints, state, local, and tribal policymakers are committed to leveraging data more effectively. State and local policymakers gave several examples of how increased data-sharing capacity would further their goals, such as helping them assess how their jurisdiction is doing relative to others, identify approaches that worked well in comparable places that may successfully translate, and identify common data and research needs. Regional data-sharing could also enable state and local policymakers to track outcomes for people or children who participated in education or programs in one state but then moved to another. Longer-term outcomes, such as post-secondary college attainment, employment, and wage outcomes, would enable more comprehensive program evaluation and improvements.



Some of the resource challenges faced by state, local, and tribal governments could be addressed by federal agencies rationalizing the data they collect through government reporting mechanisms. The federal government collects vast amounts of information on how programs such as block grants, Medicaid, TANF, SNAP, and others are administered. These aid programs are run by different agencies but often include the same participants, resulting in the same data being reported multiple times to the federal government. Further communication and collaboration between federal, state, tribal, and local governments could help ensure that scarce resources are not wasted collecting suboptimal and redundant data.

Researcher User Needs

Foundational research questions about what drives economic mobility remain unanswered and are critical to understanding the full range of approaches necessary to reverse the multi-generational decline in economic mobility. Researchers perform many tasks that inform policy and the public, providing timely analysis of present conditions, evaluating programs or policies for impact and effectiveness, and studying outcomes over years or generations, among many others. Across this broad range of approaches and methodologies, researchers agreed that data were difficult to discover, were non-standardized, and lacked key descriptors regarding potential bias or gaps in government data. Government processes to access data are often lengthy and burdensome, and researchers shared that it was often difficult to identify opportunities to collaborate with other researchers who had worked with the same data. Researchers also struggled with data timeliness, which presented challenges for researchers and analysts who worked directly for or with policymakers.

o Data Discoverability, Access & Descriptors

Increasing access to standardized, linked, privacy-protected data-sharing for both governmental and external researchers and data scientists facilitates program and impact evaluation that can further inform effective policymaking. While some research institutes or universities have developed effective partnerships with federal and state agencies for ongoing data access, many researchers must navigate cumbersome approval processes to access government data. Even researchers within government agencies or legislative bodies can experience years-long delays in trying to access data from other agencies. Researchers are also often stymied by the lack of data discoverability, despite nascent efforts by the federal government to require agencies to create public data inventories and a pilot project with the University of Michigan to create a single sign-on for researchers to apply for data access. Even if data become more easily discoverable through these efforts, the back-end challenges of having to work through agency-by-agency approval processes will remain unless they are directly addressed.

Beyond accessibility, government data are often not standardized, particularly at the state and local level, and data collected from multiple sources may have a variety of meanings. This, in turn, makes the task of aggregation and statistical analysis challenging. When there are data descriptions, they rarely alert users to potential bias in individual data sets or bias or gaps that may emerge when data sets are combined. Bias, gaps, or underrepresentation can create or compound the risk of inaccurate or biased outcomes when those data are used for predictive modeling or automated decision-making. Better descriptors about how data have been used in past research; clear assessments of bias, undercounts, or gaps; and technical approaches to data cleaning that mitigate bias would greatly help researchers understand the pros and cons of using different data. Publicly available, comprehensive data descriptors will also allow for external assessment of research created from those data, increasing research integrity and replicability.



Timely Data Analysis to Inform Policy

Researchers and data analysts, including those embedded in government agencies, need more real-time data to inform policymaking decisions more rapidly. This is especially critical in times of economic crisis, when lags in economic data on income, unemployment, housing stability, and food security make it difficult to understand rapidly changing needs and to craft responsive policies. As discussed in more detail below, there is increasing experimentation with private sector data to reduce the time lags, though challenges remain in the comprehensiveness and potential bias of the data, such as coverage in underbanked and marginalized populations and a lack of data on race, ethnicity, and gender in many current iterations. Exploring combinations of non-traditional data—such as private-sector and nonprofit data—with federal data sets may provide an avenue to increase timeliness, but this strategy should be evaluated for risks prior to government adoption, including evaluating transparency, bias, and the ability to cover marginalized populations comprehensively.

• Public User Needs: Increased Education and Engagement

Data can provide valuable insights to individuals, communities, and businesses as they make decisions about education, careers, and where to live or invest in order to increase stability and opportunity. Maximizing access to open, machine-readable government data, dashboards, and other data tools promotes government transparency and accountability, innovation, and civic engagement and could help increase public trust in government. The experts cautioned, however, that a long history of government data used for enforcement and in other harmful ways may contribute to heightened distrust, particularly in communities of color and marginalized groups.

To increase public awareness and address data equity, the federal government should engage in a largescale, concerted effort to help the public better understand data collection and the benefits that data use can provide. Increased engagement should position communities as experts in their own needs, contribute to the creation of diverse data use-cases, and create additional valuable external demand signals to inform data improvements. However, the experts noted a lack of existing best practices or research on the most effective way to engage with the public or marginalized communities on data practices. There are promising emerging practices using participatory research and even participatory algorithms which engage affected communities throughout the research or development process and which could inform data-focused engagement. Documenting and evaluating different approaches to public engagement could do a great deal to help governments, advocates, nonprofit organizations, and others undertake the increased public engagement the experts see as central to increasing data equity.



PART TWO: CHALLENGES AND OPPORTUNITIES IN THE CURRENT DATA LANDSCAPE

In addition to understanding the high-level needs of key end-users of government data, the expert consultations revealed a broader landscape of challenges and opportunities in government data that we used to inform our strategic roadmap. We share the key takeaways from the expert engagement below, focusing on opportunities to increase capacity at the federal, state, and local levels; to increase data equity; to improve public and community data; to create external demand signals; and to support experimentation and innovation in non-traditional data and measures.

Improving the Quality and Accessibility of Federal Data

The federal government's statistical and administrative data collection is by far the most robust and the most geographically and demographically comprehensive data set available in the United States. Federal data—especially in combination with state administrative data and other non-traditional data—offer the greatest opportunity to answer fundamental questions about how to accelerate economic mobility for all. Increasing the federal government's capacity to link data sets across agencies, across levels of government, and with other external data was identified as a top priority. We discuss the key expert takeaways on federal data, share information on existing efforts, and share a range of strategies the experts identified for advancing the quality and accessibility of federal data.

Key Takeaways

• Federal reforms will require long-term effort.

In the past few years, the federal government has taken several promising steps to improve data processes. The ongoing implementation of the <u>Foundations for Evidence-Based Policymaking Act</u> and the <u>Federal Data Strategy</u> offer helpful frameworks to improve the transparency of federal data assets, requiring all agencies to create and organize their data efforts around learning agendas, to appoint a Chief Data Officer, and to create a single-sign-on portal for researchers to request data access from statistical agencies. However, to date, substantive decisions on prioritization of efforts have been left to individual agencies, implementation varies widely, and there is no effort to identify cross-agency issues which may require coordination. While progress has been made, many agencies still operate with a "need to know," not a "need to share," attitude towards expanding data access.

Several of the Biden administration's earliest executive actions or orders include major data initiatives and are promising indicators that the pace of internal federal data reforms will accelerate. The first full week of the administration saw the establishment of an Equitable Data Working Group, co-chaired by the U.S. Chief Statistician and the U.S. Chief Technology Officer and including the Director of the Office of Management and Budget (OMB) and key data agencies such as the Treasury Department and the Department of Commerce, as well as senior White House policy office leaders including the Chair of the Council of Economic Advisors, the U.S. Chief Data Scientist, the U.S. Chief Information Officer, and the Administrator of the U.S. Digital Services— all of whom could help implement policy recommendations that may emerge. Among other responsibilities, the Equitable Data Working Group is tasked with assessing opportunities to improve the disaggregation of data by race, ethnicity, gender, disability, veteran status, and other key demographics.

The <u>Memorandum on Restoring Trust in Government Through Scientific Integrity and Evidence-Based</u> <u>Policymaking</u> similarly tasked OMB with creating guidance to inform agency evidence-building and annual evaluation plans, directing all agencies to expand access to federal data to support evidencebuilding, to make data available in machine-readable formats, to increase access for researchers and the public, and to publish an agency data plan that provides a consistent framework for data stewardship, use, and access. The memorandum also empowers the Council of Chief Data Officers to establish government-wide best practices for the use, protection, dissemination, and generation of data that incorporates scientific integrity. In short, many of these early executive actions will begin to address several of the key challenges with federal agency data processes identified by the experts—and strongly suggest that the executive branch may be a willing partner in the strategies identified below.

While there are near-term opportunities for targeted reforms that can increase the quality and accessibility of federal data, comprehensive federal reform will likely take years and require sustained leadership from the Office of Management and Budget and other senior executives in the White House. Indeed, the Committee on National Statistics at the National Academy of Sciences estimates that necessary reforms to improve federal statistics will take a decade or more. Combining the near-term opportunity of a potentially supportive administration with the reality of the longer-term nature of federal reforms, the experts identified strategic opportunities to increase government capacity to accelerate internal reforms and external investments to help inform and focus federal efforts on the highest-value data and processes.

• The federal government needs increased data-linkage and sharing capacity.

Leading experts, including the Committee on National Statistics (CNSTAT), have written extensively on the need for increased federal capacity to link and share federal data, leveraging leading-edge privacy practices and ideally including state and local administrative data, while exploring opportunities to include other non-traditional data. The final report of the U.S. Commission on Evidence-Based Policymaking included a recommendation to establish a new National Secure Data Service (NSDS) to meet this need, a recommendation that was not included in the Evidence Act that passed in 2018. Recognizing the ongoing need for increased, privacy-protected data linkage and sharing, data experts have continued to explore the most promising avenues to establish an NSDS. A recent paper co-authored by former U.S. Chief Statistician Nancy Potok and Nick Hart, President of the Data Foundation, provides an in-depth analysis of the core components and capacities necessary for a successful NSDS, outlines different approaches to achieve sustainable funding, and examines the pros and cons of different options to house the new entity. The paper ultimately recommends the creation of a new Federally Funded Research and Development Center (FFRDC) to house the NSDS, leveraging existing authorities at the National Science Foundation's National Center for Science and Engineering Statistics (NCSES). The NCSES qualifies as a federal statistical agency and is covered by a core privacy law governing the use of statistical data, the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), that would enable the new entity to link and share federal statistical data without requiring new legislation. The experts were encouraged by the level of consensus on the core capacities needed in an NSDS and were optimistic that the new administration may establish it. Recognizing that funding and standing up a new entity, hiring skilled personnel, implementing a technical infrastructure, and creating data-sharing agreements will take time, the experts also recommended additional near-term improvements to advance federal data access and sharing.

• The U.S. Census Bureau should reduce barriers to data access.

While all federal agencies have valuable data on government programs and policies, the U.S. Census Bureau (Census) has the most centralized collection of valuable longitudinal data for research in the country. It has comprehensive race, ethnicity, age, gender, income, marriage, death, home ownership, education, disability, and location data over several decades that, if made more accessible, could dramatically lower the costs of rigorous research looking at the long-term outcomes of government interventions and programs. Census is also working on a project, <u>the American Opportunity Study</u>, to digitize paper records from past decennial Censuses, which could eventually extend historical data back to the 1960s, allowing for unprecedented research across multiple generations.



With its rich, individual-level demographic data, Census could also play a central role in improving federal data equity by linking or appending federal or state data sets to add race, ethnicity, gender, and other demographic fields to data sets where they are lacking. Increasing the federal government's ability to disaggregate data by key demographics could significantly expand opportunities to understand the impact of programs intended to provide stability and mobility across populations, geographies, income levels, and other factors. Despite these rich and growing data assets, the limitations placed on Census's data approval processes—both by provisions in its authorizing legislation, Title 13 of the U.S. Code, and by the Bureau's own business practices—were frequently identified by the experts as presenting challenges for researchers and policymakers.

A number of agencies have shared valuable data on the economy, health, employment, and income/wages, including the Department of Veterans Affairs, the Social Security Administration, the Internal Revenue Service (IRS), the Bureau of Labor Statistics, the Bureau of Economic Analysis, the Department of Housing and Urban Development, and the National Center for Health Statistics, among others, and Census is continuing to add additional agency data to its collection. To support these efforts, the <u>Evidence Act</u> included language that required other agencies to share their data with OMB-designated statistical agencies upon request, with some narrow exceptions. Despite that affirmative language, the experts observed that, to date, the pace of agency data-sharing with Census (and other statistical agencies) has remained slow.

We also heard from policymakers and researchers that Census data access and review processes are lengthy and burdensome, particularly when research questions require access to data from other federal agencies housed at Census. Each agency that has shared data with Census has negotiated individual, one-to-one memoranda of understanding to govern access to their data, many of which require researchers to apply separately to each agency's research approval process. Once data access is granted, Census and the other agencies also have separate processes to review research prior to publication. These processes are time-consuming and inefficient and may disadvantage researchers early in their careers or at less-resourced institutions.

Another limiting factor is that, until the COVID-19 pandemic required remote virtual access, almost all research conducted with Census microdata occurred in brick-and-mortar Federal Statistical Research Data Centers (FSRDCs). There are currently thirty FSRDCs around the country, affiliated with fifty research organizations, universities, federal reserves, and nonprofit organizations. The relatively limited number of FSRDCs presents obvious limitations for researchers who are not nearby or affiliated with related universities or organizations. During COVID-19, many FSRDCs had to shut down or provide limited access. Census began to show some flexibility on remote access during COVID that could hopefully translate into a more robust, intentional approach to increase secure, online data access. However, the current Census approach, which requires researchers to use the Census Bureau-provided Virtual Desktop Infrastructure, would be difficult to scale up for broader national access. Several experts have advocated for Census to move to sharing data via secure, cloud-based servers to expand access, pointing out that the modern privacy and security standard required of federally approved cloud service providers could provide the same, if not enhanced, protections for Census data. The U.S. Census Bureau funded a cloud-based data access infrastructure pilot project at New York University that has developed into a successful independent research data access platform, but it has not followed up on this work.

Rigorous privacy protections are compatible with increased data sharing, but trust in federal processes has eroded.

Ensuring that data collected from individuals, communities, and businesses are rigorously protected is critical to ensuring public trust and was a central point of discussion across all the expert engagements. The experts started with a recognition that different users will require different levels of data access,

ranging from open data to highly restricted individual-level data. and that clearly articulating different needs across different data use-cases would be a valuable step in increasing transparency about privacy and data access. The experts recommended establishing data privacy and governance frameworks that determine appropriate privacy protections based on use-case and user need. Preliminary questions to guide a privacy framework could include:

- What will the data be used for and who is the primary user?
- What is the highest level of data aggregation that will meet the specific needs of the user?
- How do we address data repurposing?

MARKLE

• How much information or control should be provided to people about how their data may be used?

For example, while policymakers need data that is timely and geographically granular, they will almost never require individual-level data. Researchers may need to track individuals across interventions or over time but will rarely need to know personally identifiable information (PII). Secure data-linking processes can substitute numeric or other identifiers for PII to assure researchers they are tracking the same person without ever seeing or publishing individually identifiable information. The Census Bureau uses this approach to link external data with Census data, creating a Protected Identification Key for researchers accessing microdata and providing match rates so researchers know how much of their original data was successfully linked. In addition to privacy protections, as discussed elsewhere in this paper, researchers need ongoing education on ethical and responsible data use, including understanding how bias or gaps in data may raise equity issues in research conclusions or in the use of automated decision-making tools.

The Evidence Act requires OMB to issue a regulation on how federal agencies should classify data according to its sensitivity and provide access accordingly. This regulation is currently overdue, but once it is issued, it should help agencies more systematically to determine how to provide access to data for researchers. As part of the current action items included in the federal data strategy, a cross-agency team is developing a data privacy toolkit for agencies to use in considering how to protect the identities of individuals and businesses whose sensitive information is being shared or linked. The federal government continues to experiment with new privacy protection and disclosure techniques, such as differential privacy and secure multi-party computation, to provide researchers with granular data access to facilitate rigorous research while preventing reidentification or other privacy harms.

Transparency and clear explanations of privacy protections, data use, and access protections must be a core part of government privacy and governance frameworks, especially amidst concerns about the erosion of public trust in government data process. The experts pointed to lack of public understanding about how individual data is used or reused by the government and how existing protections protect privacy, as well as heightened concerns that the federal government will use routinely collected data from individuals, businesses, and communities for punitive purposes. The federal government should determine the minimal amount of data that needs to be collected to meet needs, while ensuring that those data are maximally responsive to community needs. Ideally, the federal government will establish clear guidelines preventing enforcement agencies from accessing administrative or statistical data, with narrow, clearly defined exceptions, such as maintaining program integrity and preventing fraud or abuse.

• Effective communication around the risk of data misuse should also emphasize the benefits of privacy-protected data sharing.

A key part of data responsibility is not just preventing misuse, but also preventing "missed use." Much of the focus of data sharing and use has emphasized the need for privacy protections to prevent misuse that harms individuals, groups, or businesses. While privacy protections and preventing misuse are critical, participants see significant opportunity in shifting the conversation from a defensive posture - which may reinforce perceptions of risk to an offensive posture emphasizing the benefits of data collection and use. A comparison was made to people choosing to accept data tools and services, e.g., smartphones and apps, where they agree to data sharing because they see value in the tool they access.

In addition to ensuring strong privacy protections, the federal government can also focus communications on emphasizing the benefits of government data use, informed by community engagement and input on use-cases, as a starting point, coupled with ongoing rigor to ensure and communicate privacy protections. For example, New York City placed subway and bus stop ads showing particular services with the tagline "Made by Open data" to provide compelling, concrete examples of beneficial data use. Some experts pointed to a similar model used by the Public Broadcasting Service which opens programs with the tagline "paid for by viewers like you" to engage viewers as co-creators of content, while also providing assurances that "no animals were harmed in the making of this program" as an example for consideration. Applying this "sandwich" approach of leading with the positive benefits of data, while providing assurances that negative outcomes are guarded against, could present an interesting communications model for the federal government to replicate.

• Federal and state agencies need consistent, predictable funding to improve data collection and sharing.

The experts had many suggestions of ways to leverage OMB to accelerate implementation of federal data strategies, ensure cross-agency collaboration, and provide effective oversight and accountability to ensure progress and compliance. One innovative suggestion was for OMB or other White House policy council processes to create a "gating" policy, requiring all policy proposals to state specifically what data they used to formulate the policy, how they will use data to track policy implementation, and what data gaps may exist. A related suggestion was to establish a nonpartisan external Data Council, similar to Australia's Productivity Commission, an independent government agency which assesses proposed policies to determine whether sufficient data exists to support the policy, its implementation, and to enable evaluation of policy impact. Another suggestion, outlined in an <u>article published by the National Academy of Public Administrators</u>, proposes a reorganization of OMB to establish a new position, Assistant Director for Information Policy, to oversee, manage, and coordinate data activities across OMB's divisions and offices.

In addition to these internal reform ideas, the experts were clear on one core need: they agreed that the most important component to ensure ongoing implementation of government data improvements was predictable, long-term budget support to agencies to increase their data capacity. Establishing budget set-asides to support data infrastructure improvements has a demonstrated track record of driving data improvements in the past at both the federal and state levels. For example, the American Recovery & Reinvestment Act of 2009 provided a 0.5% set-aside which could be used to support data infrastructure improvements in the Act of 2009 provided a 0.5% set-aside which could be used to support data infrastructure improvements to meet the increased reporting requirements in the Act. Similarly, as a part of the implementation of the Affordable Care Act, the federal government covered up to 90% of the cost of building new Medicaid enrollment systems and up to 50% of the cost of including enrollment for other federal benefits, such as the Supplemental Nutrition Assistance Program (SNAP) or Temporary Assistance for Needy Families (TNAF). Direct federal funding was also highly successful in incentivizing states to create State Longitudinal Data Systems, with all fifty states, the District of Columbia, and Puerto Rico having the ability to link data between education, early childhood, and workforce data systems.

MARKLE



• Strategies to improve the accessibility and quality of federal data.

The experts raised a range of strategies and opportunities to address these challenges, focused on;

- 1. Increasing the federal government's internal capacity and providing additional expertise;
- 2. Reducing barriers to accessing linked, longitudinal federal data;
- 3. Creating strong external demand signals to help establish an effective feedback loop between key stakeholders and federal agency data supply efforts; and
- 4. Expanding and diversifying research leveraging federal data while democratizing access.
- Proposed Strategy: Increase Expert Capacity in Federal Government to Advance Federal Data Reforms:

There is good reason to believe that the new Biden-Harris administration will be supportive of existing data strategies, new strategies to increase data disaggregation, and increasing access to and uptake of safety-net benefits. The Administration has created a new role focused on technology, service, and delivery in White House Policy Councils and in key agencies. However, given the magnitude of crises the new Administration is tackling, increased expert capacity could provide the support needed to advance efforts to improve the quality and accessibility of federal data.

Leveraging <u>Intergovernmental Personnel Agreements</u> (IPAs) could help swiftly onboard experts in data equity, privacy, linkage, and external engagement who could accelerate federal data improvements and provide needed skills that may be hard to acquire through traditional government hiring. IPAs can come from nonprofit organizations, which can include experts from state and local governments, tribal governments, universities, research and development centers, and certain nonprofit organizations such as the Partnership for Public Service. IPAs are an underused resource and could help increase expertise and capacity on the following priorities identified by the experts, along with others that may arise:

- Improve federal data linkage and sharing processes, including accelerating partnerships with state and local governments, researchers, and the public, particularly if a National Secure Data Service is established;
- Serve as "data stewards" or in a customer service role in key federal agencies to increase responsiveness to external stakeholders and help translate external demand to internal data processes;
- Undertake a data equity audit of federal data and document opportunities to identify and fill gaps in data disaggregated by race, ethnicity, gender, and other key demographics;
- Support the creation of OMB guidance on Evidence Act implementation to accelerate agency development of learning agendas, data inventories, and improving access to data;
- Identify opportunities to leverage federal funding through budget set-asides or other consistent funding streams, in order to support federal agencies in improving their capacity to collect and share data and increase data disaggregation;
- Develop and implement agency engagement strategies for more effective outreach to diverse communities, in order to increase input, better understand community data needs, and provide education on the benefits, uses, and protections in federal data; and
- Identify metrics and data to track the implementation and impact of stimulus or emergency funding to inform future policymaking.



• Proposed Strategy: Reduce barriers to linked, longitudinal federal data.

As discussed above, establishing a National Secure Data Service (NSDS), or a similar entity, to increase the privacy-protected sharing of linked data could help address many of the challenges of federal data access, protection, and sharing. Given the amount of excellent work already published on the capacities, location, and scope of an NSDS, we focus our strategies on opportunities to improve existing data processes to streamline access and reduce burdens. Creating an overarching data governance framework to guide consistent data access across federal agencies could provide needed clarity and consistency and could make future data-sharing agreements easier to negotiate. Short of a formal governance process, the experts identified several high-leverage opportunities to improve the U.S. Census Bureau data processes that could have outsized impact, given the breadth and range of the Census data assets.

Additionally, the federal Interagency Council on Statistical Policy, chaired by the U.S. Chief Statistician, established a governance steering committee to guide expansion of the Federal Statistical Research Data Centers to provide more access for researchers and to expand data sets available through this network. The steering committee consists of a subgroup of operators of the data centers and heads of the statistical agencies. This group would guide many of the changes recommended for Census, since the vast majority of Census microdata is accessed through the Federal Statistical Research Data Centers, which Census manages on behalf of the federal agencies making their data available through that mechanism. The steering committee should be key in addressing the recommendations, and since it has no funding and very little staff support, it could also benefit from IPAs. The suggested improvements to Census processes listed below could be advanced through IPAs at Census or other data agencies or proposed through the federal <u>Advisory Committee on Data for Evidence Building</u>. Specific recommended improvements include:

- Census and agencies whose data it holds could create a unified research approval and review process, eliminating the need for separate applications;
- Census and other data-holding federal agencies could affirmatively identify common data use-cases or categories of research questions for which, if identified criteria are met, approval could be automatic or significantly fast-tracked;
- Census could create expedited access for other government entities, e.g., state and local governments who seek to link their administrative data with demographic and wage data or government economists seeking data to inform legislation or budget decisions;
- Census should move away from requiring access approval on a project-by-project basis and provide approval for groups broadly seeking access to data for similar or related research questions; and
- Accommodations or access permissions that are made for one group could automatically apply to other similarly situated groups..

• Proposed Strategy: Create strong external demand signals.

Providing federal agencies with coordinated, clearly articulated demand signals from key stakeholders will significantly facilitate federal government responsiveness by clarifying where to turn to get input on data, policy, and research questions. Externally organized groups or coalitions could work across their constituencies to identify and vet data and research needs, create priorities by consensus, and create agency engagement strategies. Ideally, once initial engagement has been established, federal agencies would maintain ongoing communication, building increasingly effective feedback loops between demand and supply that can adjust over time as priorities change. Coordinated vetting and prioritization processes from key stakeholders would also benefit the broader data ecosystem, helping focus funders, researchers, entrepreneurs, and others on areas of need or opportunity. Efforts to organize external

demand could be organized and supported in a number of different ways, with the experts identifying several different options, including:

- Support for policymaking stakeholder organizations—such as the National Governors Association, National Congress of American Indians, National Association of Counties, National League of Cities, and the U.S. Conference of Mayors—to identify data, data products, and data processes that would best support state, local, and tribal government policymaking to increase mobility and improve outcomes for students and workers;
- Support for state regional data collaboratives or other state coalitions to engage with federal agencies on access to enhanced data to improve education, labor-market, and workforce training policymaking, and to contribute to federal research prioritization; and
- Support to establish a nonpartisan external Data Council, similar to Australia's Productivity Commission, to assess whether there is sufficient data to evaluate and ultimately execute proposed policies, to bring transparency and clarity to such decisions, and to enable evaluation of impact.
- Proposed Strategy: Expand research leveraging federal data and democratize access:

Challenges in accessing government data limit the number of researchers who can study key questions concerning economic stability and well-being. Increasing and diversifying the number of researchers and research questions asked of federal data will provide valuable new insights into stability, mobility, and economic well-being. We also anticipate that supporting expanded research will advance data equity by increasing data disaggregation and encouraging investigations into the most effective ways to engage with the public and especially with marginalized communities. Expanded funding for researchers directly leveraging federal data will also help to telegraph more broadly the value of federal data for research and to identify data sets that are most requested and how they are used. Funders would be able to identify high-priority research questions that would come up through the efforts described above, or they could create open funding pools with the purpose of identifying novel, diverse, and/or innovative research questions.

Wherever possible, funders should require that researchers create public goods from the data they have leveraged, further democratizing access to federal data. While public data goods are not equivalent to individual-level public data access (appropriately, due to privacy protections), they can still enable more fine-grain analysis than is currently available. As part of the creation of public data goods, we recommend that funders require researchers to include detailed data descriptors when they publish research, including any identified bias or gaps, to help supplement and explicate existing data.

Suggestions for research funding support that would best advance federal efforts in the nearer term include:

- Create a pool of research funding with the explicit goal of increasing the number and diversity of researchers who access federal data;
- Support additional research hubs or institutes to access Census and other federal data to make it easier for a broader range of affiliated researchers to access federal data; and
- Support the creation of more "public data goods" like the <u>Opportunity Atlas</u> which provide broad public access to federal or private sector data.
- Proposed Strategy: Continue progress on sharing open, machine-readable data.

Open data has been one of the most universally adopted approaches to increasing transparency and accessibility of data. Open data, public data products, and transparency around federal privacy protections will also help public engagement efforts and will allow a broader range of individuals,

MARKLE



communities, and businesses to leverage data to answer questions, making more informed decisions. Full implementation of the Federal Data Strategy and the Evidence Act open data requirements should be implemented and expanded, including:

- Creation and publication of Open Data plans, with robust opportunity for external input;
- Creation and publication of agency data inventories and data catalogues that include data descriptors to inform users of gaps, bias, or other potential challenges in the use of data sets;
- Creation and publication of data governance and accountability infrastructure; and
- Ongoing OMB guidance and oversight to ensure agency compliance with open data requirements.

INCREASING THE ACCESSIBILITY AND USE OF STATE, LOCAL, AND TRIBAL DATA

States play a unique role in the data ecosystem, often serving as an intermediary between federal and local governments, overseeing the delivery of many federal safety-net benefit programs, collecting a wide range of data, and procuring and maintaining core data and technology infrastructure. Local and tribal governments often serve as the formal or informal "last mile" of program delivery, bring valuable insights on diverse community needs, gaps, and opportunities to improve delivery. Many of the experts we spoke with recognized the powerful insights that could be gained by creating an effective ongoing data sharing feedback loop between the federal, state, and local government. However, much like their federal counterparts, state, local, and tribal governments often lack the internal expertise to effectuate broader data-sharing. Despite their central role in oversight, administration, and delivery of core mobility programs, state, local and tribal governments often struggle with out-of-date technology and data infrastructure that make implementing policy changes or sharing data difficult. Major improvements can be made by increasing capacity at the state, local, and tribal level, sharing enhanced federal data back to states, and increasing exploration and understanding of how benefits and services are delivered.

The COVID-19 crisis provided a stark reminder that America's social safety net provides critical support in times of economic shock—but that aging technology and data systems too often get in the way of effective benefits delivery. These challenges manifested most notably in unemployment benefits systems, where multiple states struggled to provide enhanced benefits and to enroll new beneficiaries, leading to significant delays for thousands of newly eligible unemployed workers. States similarly struggled to quickly implement the new pandemic Electronic Benefits Program intended to provide direct funding for food support to families with children who were eligible for free and reduced lunches but who were no longer in schools, even as food insecurity increased.

The pandemic also demonstrated the need for better demographic data, as many state and local governments did not collect race or ethnicity data on COVID cases and deaths, impacting insights on the disparate impact on communities of color, which was readily apparent but not easily quantifiable due to the lack of hard data. The lack of data on the demographics and location of people who access key safety net benefits, such as the Supplemental Nutrition Assistance Program (SNAP) and Temporary Assistance to Needy Families (TANF), made it more difficult to identify the growing population of people who were eligible for benefits, whether they were applying for those benefits, and why some who were eligible were not applying. Without these key metrics, governments did not know how to increase benefit uptake for people in need, could not understand how uptake varied across populations or geographies, and could not assess the impacts of access to benefits on stability and health in these emergency circumstances. This lack of data, combined with the challenges in state data infrastructure, limits the ability of policymakers, researchers, and the public to increase equity and improve outcomes for people in need of support.

MARKLE



Key Takeaways

• State government data collaborations offer great promise.

States are increasingly coordinating and sharing data with each other, with a growing number of states joining regional data collaboratives. The first such collaborative is a collection of nine midwestern states, many of which leverage the <u>Administrative Data Research Facility (ADRF)</u>, a nonprofit sharing platform to share data with each other, identify common data analysis needs, and create collaborative data products to meet those needs. In addition to sharing data across states, the Midwest Collaborative also plays a key "middleman role" by engaging with local governments and workforce boards to identify their data and research needs, then working across states in the collaborative to vet and prioritize research questions and data needs. This research, data, and policy vetting function could be used to identify external demand, providing federal agencies to seek enhanced data to meet the identified demand. There are active conversations underway to create more regional data collaboratives with northeastern and western states. Currently, these collaboratives are focused primarily on workforce training and labor market conditions, though they also explore post-secondary data, and could likely expand to other priority issues areas.

State and local governments are also working to build internal capacity with emerging communities of practice such as the <u>Network of State Chief Data Officers</u>, to share best practices, and to prioritize usecases. From 2016-2017, the Ash Center at Harvard ran a similar group for city CDOs, the <u>Civic</u> <u>Analytics Network</u>. Additionally, multiple cities have leveraged the <u>What Works Cities</u> network to increase data sharing. However, in many state agencies, the only dedicated full-time data staff are focused on meeting federal data reporting requirements, and many cities and counties, especially the smaller jurisdictions, have no dedicated data capacity. While these efforts are a good start, experts agree that increased data infrastructure and analysis capacity is still a significant need for most state and local governments.

• State administrative data shared with the federal government should be enhanced with federal data & returned to improve state, local, and tribal insights.

Providing states with their own data enhanced by federal data would support more targeted state policymaking by population and geography, improve benefits delivery, and allow for better prediction of labor markets, training, and education needs. The federal government has data that is unavailable to the states, including data on wages, location, detailed information about people's employment, health outcomes, and crime victimization. The federal government also has demographic data that could fill gaps and address disparities in state data, as well as enabling national coverage that would allow states to see the outcomes of their programs on people who move across state lines. Key labor market data regarding unemployment are aggregated up to the state level, and information on where people work is clustered by industry, often making it difficult for state and local governments to understand with any geographic specificity where layoffs are occurring and where retraining resources should be focused. In one instance we reviewed, state data showed an unemployment rate of 10% for the mining industry at the aggregated state level, but additional data showed that in several adjacent counties on the state's border, mining industry layoffs comprised over 40% of county-level unemployment. With that level of information, the state and local governments could deploy more targeted retraining support.

• Data on Benefits Delivery are Critical in Response to Economic Shocks

In the current economic crisis—and in the immediate future—state and local governments need more information to improve benefits delivery, including understanding who is eligible but not accessing benefits, who has applied for benefits but not received them and why, and how to increase uptake.



Increased data linkage and sharing offer new opportunities for research on how government interventions, either individually or in interaction with each other, impede or advance mobility across geographies, populations, and varying levels of government support. In addition to identifying who is falling through the cracks, we also need to understand the implications of not receiving support. We hope to support increased stability and mobility for at-need populations in the current crisis and beyond by improving state and local data infrastructure and increasing their capacity to leverage data to improve service and benefits delivery.

• State reporting requirements should move beyond a primary focus on compliance.

The experts also highlighted the challenge of meeting federal reporting requirements, which can be overly focused on technical compliance with program requirements, while neglecting to consider data on key outcomes which could help states better understand and improve their programs. Due to this disconnect, limited resources are squandered to produce data and reports that provide little to no insight on population needs and program effectiveness. Especially in states with small budgets, funds could be allocated more productively. The experts we consulted recommended a shift in reporting requirements to prioritize data that will help both the federal and state governments improve program effectiveness—data such as who is accessing the program, whether it is achieving its intended outcomes, and what improvements may be needed, in addition to ensuring appropriate fiscal compliance.

Strategic Opportunities at the State, Local, and Tribal Level

Given the longer-term timeframe and sheer scope of implementing widespread federal data reforms, investments at the state level offer great near-term potential. The experts see great promise in linking and sharing state administrative data to support more effective policymaking and research, especially if enhanced with federal data. However, despite some advances in modernizing data infrastructure, often spurred by federal funding, many key state data systems are outdated and unwieldy, as revealed when states tried to administer new unemployment benefits quickly in response to the COVID-19 crisis. We see significant opportunities in state-level investments to increase data capacity, support the trend towards regional collaboration, and expand the number of state-level linked, longitudinal data sets to support better policymaking and research. States could play a valuable role as the "connective tissue" in a federal-state-local data feedback loop, discovering and vetting priority data and research questions from local governments to make clear to the federal government that there is a nationwide demand for high-quality, detailed data.

While most of the experts' suggestions focused on state-level opportunities, we also engaged with stakeholder organizations and experts who work closely with local governments. Local and tribal governments also struggle with internal capacity, with the need particularly acute in smaller, more rural, or tribal governments. They often do not have the budget to hire expert staff and may not have sufficient data to interest a university partner. Some of their needs could be met by public data products that were more responsive to local government needs, and many of them already used Census data offerings to inform local decisions. Ongoing engagement with local and tribal governments will help identify better public engagement methodologies, improve service and benefits delivery to marginalized and hard-to-reach populations, and increase equities—all of which should be critical priorities in state and federal efforts.

• Proposed Strategy: Increase capacity and support expansion of regional data collaboratives.

The state collaboratives are currently run by their senior agency leaders, inherently limiting the amount they can accomplish while still running their agencies. Increasing engagement with federal agencies offers multiple potential benefits, positioning the collaboratives (including individual states where collaboratives don't currently exist) as coordinated sources of demand, identifying data and research



priorities from local governments, articulating common needs across regions, and, ideally, making it easier and faster to negotiate data-sharing agreements with federal agencies that will increase the flow of state administrative data to the federal government while also increasing the amount of enhanced federal data provided back to state agencies. Suggested investments to support collaboratives include:

- Investments in staff support such as project managers, data scientists, and engagement experts;
- Support for collaborative convenings to surface and share best practices;
- Support for secure, privacy-protected data-sharing across collaborative partners and with federal agencies; and
- Support for data product development and deployment across participating states, including local and tribal government priorities, and including open/publicly shared data.
- Proposed Strategy: Increase state government data capacity.

As states seek to improve their data infrastructures to improve benefits and program delivery and leverage the data generated from those systems, there is value in identifying and providing funding support for vetted technology platforms for data linkage, and sharing could provide added capacity and increase innovation. Additionally, supporting internal data expertise could help create data governance processes to increase data sharing, increase analysis and product development, and ensure privacy and security protections. Potential opportunities to increase state data capacity include:

- Providing data analytics capacity to state agencies to help them establish or advance within-state data sharing governance and processes, increase linked, longitudinal data, including data from local and tribal governments, to inform policymaking and share with researchers.
- Create a process to identify and vet technology platforms to meet state needs for data linkage and sharing, as well as benefits application and service delivery systems. Multiple viable and vetted providers could provide needed capacity and increase innovation.
- Provide a pool of funding to support new entrants into the state and local market that meet certain requirements such as interoperability, privacy, and security, leveraging agile design, and reasonable price points;
- Support university-based partnerships to provide ongoing data support such as cleaning and linking of data, managing external data access requests, and doing independent research, with government research priorities receiving precedence.

• Proposed Strategy: Increase research and analysis focused on improving benefits delivery and evaluating program impact.

In the midst of the current economic crisis—and looking to the near future—state, local, and tribal governments need more information to improve benefits delivery, including understanding who is eligible but not accessing benefits, who has applied for benefits but not received them and why, and how to increase uptake. Increased data linkage and sharing offer new opportunities for research on how government interventions, either individually or in interaction with each other, impede or advance mobility across geographies, populations, and varying levels of government support. Opportunities to increase insights on how these key programs are delivered and how they impact mobility include:

• Support for technical infrastructure improvements to modernize benefits application and delivery systems, including integrated benefits systems;

- Support for increased research and analysis of benefits applications processes, including how to do outreach most effectively to eligible beneficiaries, how to increase benefits uptake, and to better understand the "pain points" throughout the process where improvements could be made while maintaining program integrity; and
- Support for expanded research on how state programs improve stability and mobility, both individually and in combination.

INCREASING EQUITY IN GOVERNMENT DATA IS CRITICAL TO ADDRESSING DISPARITY AND INEQUITIES

Public trust that the government's collection and use of data is fair, equitable, and used to improve people's lives is necessary to promote racial justice and ensure public legitimacy. Much work is required to build public trust, particularly at the federal level, where increased data disaggregation, meaningful community engagement, and public education could have the greatest impact. Supporting expanded research to assess and increase current data disaggregation—and to investigate open questions about the most effective ways to engage with the public and marginalized communities—could help inform federal government efforts, particularly given the magnitude of the task and the relative lack of existing research.

Key Takeaways

MARKLE

• Increasing data disaggregation and mitigating bias in data are critical steps to increasing data equity.

Federal data and statistics are often not disaggregated by race, ethnicity, or other key demographic variables. Marginalized groups, such as undocumented immigrants, unbanked populations, tribal nations, or people in the U.S. territories, are often undercounted or underrepresented in federal data. Additionally, declining survey response rates in areas with large numbers of single-parent households, families with young children, workers with long commutes, and areas with high crime rates could exacerbate challenges to representation in federal data. Administrative data from federal and state programs can be a helpful source of additional demographic data, but are also vulnerable to inaccuracies, incomplete or missing data, and a lack of standardized definitions for different groups and inconsistent collection processes.

Administrative data often do not constitute a representative sample of a given population, due to the selection bias created by using data from individuals who have opted into program participation. Additionally, due to structural racism, minority populations are likely to be disproportionately represented in anti-poverty and benefits programs. Both under- and over-representation can create gaps and biases in existing data and may further compound racial disparities in policy development, program delivery, and outcomes. A first step towards increasing data disaggregation would be to undertake an audit or review of existing data sets to account for what is currently collected across key metrics, such as race, ethnicity, gender, and disability, assessing collection across tribal lands and U.S. territories, and documenting gaps or potential undercounts. The newly established federal Equitable Data Working Group includes a call for an initial assessment of federal data and collection along these lines and is worth monitoring to see if any of the work would benefit from supplementary external efforts.

• Statistical methods for data cleaning may introduce bias.

Experts also discussed how statistical methods commonly used in cleaning or processing data for linking, research, or use in algorithms can compound bias data. Research demonstrates that women and minorities are more likely to fail to complete demographic questions for a range of reasons, including concern that their demographic information may negatively impact the outcome of the application they are completing. When those datasets are processed for reporting, one common methodology is to



interpolate the missing demographic files, which often completes the field with the most frequent selection, thus defaulting empty fields with the majority class, such as adding male to an incomplete gender field or white to an incomplete ethnicity field. This can be particularly problematic for small population groups, because membership in these groups will never be imputed. The challenge of representation of small population groups also arises in privacy and disclosure methodologies that suppress small sample sizes to prevent reidentification. Julia Stoyanovich of New York University's <u>Center for Responsible AI</u> has identified multiple technical approaches to address these statistical challenges to reduce bias and create more accurate demographic representation in data cleaning and processing.

o Addressing data equity requires community-centered engagement.

Beyond the more technical components of data disaggregation, the experts highlighted the importance of putting the needs of communities and marginalized groups at the center of federal data efforts. While recognizing the need to engage with a diverse range of communities and highlighting emerging efforts across community organizations and local governments, the experts noted that the field is currently lacking information on the most effective ways to communicate with the public. Understanding how best to communicate the benefits and risks of data use, how data can be used to answer priority community questions, and how they may be used in critical government decision-making are key questions that require exploration. The experts pointed to methodologies being developed in the emerging field of community-based participatory action research, which actively includes members of the community whose data will be used as experts at every stage of the data and research lifecycle. Behavioral scientists in several countries have researched the best approaches to engaging with marginalized communities, particularly when conducting censuses of population and housing. Many of these best practices could easily be adapted for use in the U.S.

Some early lessons in the United States include using trusted institutions and credible intermediaries, such as schools and libraries, for community education sessions. Several local governments have undertaken community engagement efforts involving data, such as New York City's efforts to leverage libraries to do direct community engagement. An emerging approach in K-12 education is <u>"data equity walks</u>" that use data visualizations from local schools to inform parents and the public on issues of equity and disparity in schools. In addition, "citizens' assemblies" on data have been convened in <u>New York City</u> and offer an experimental way to engage the public on the topics of data re-use and regulations. The initial public conversations about COVID-19 offered a promising example of providing people with detailed information about data and its uses, empowering them to make sophisticated suggestions about data re-use and tools. Increased engagement, coupled with increased opening of data, should be prioritized as a core part of government data improvement efforts.

The experts also spoke about a more specific emerging body of work focused on government use of automated decision-making tools or algorithms that are increasingly used in applications for employment, housing, and other programs which governments directly run or fund, or for which they have oversight responsibility to ensure fairness and accountability. There are known risks of bias or harm when these tools rely on inaccurate or incomplete underlying data sets, particularly considering the known demographic data gaps and lack of representation of marginalized groups that can perpetuate systemic disparity. Awareness of the risks of these automated decision-making tools is growing, with calls for increased transparency and disclosure regarding tool use and impacts. Several state and local governments have imposed bans or required increased oversight of the government's use of these tools. Experts suggested increased public education on the use and challenges of these tools, audits to assess bias in underlying data and tool outcomes, requirements for clear disclosure to users when automated tools are used for a decision, and clear processes for end-users to request more information or dispute an automated decision.



Strategic Investment Opportunities to Increase Data Equity

Ensuring equitable and comprehensive data is critical to reducing disparities and ensuring equitable opportunity. Policymakers are increasingly eager to take steps to disaggregate their data by race, ethnicity, gender, and other key demographics, to reduce or eliminate bias in their data, to increase transparency and meaningful engagement with marginalized communities, and to establish clear guidelines that maximize privacy and prevent data misuse. In addition to ensuring that government data maximizes privacy, reduces bias, and is not used inappropriately, governments at all levels should pursue opportunities to increase privacy-protected data-sharing of existing data to better understand how programs, policies, and benefits work in combination to increase stability and opportunity. Sharing data on key programs and policies across federal agencies, state and local governments, researchers, and the public can help reduce the need for new data collection, identify where new data or measures are needed, and ensure that relevant and available data on key programs are leveraged to maximize opportunities to reduce disparities and increase racial equity. Many of the improvements to capacity, infrastructure, and sharing suggested in the strategies above would also support increased data equity. In addition to broader strategies that will support increased data equity, the experts identified additional direct strategic opportunities including:

- Proposed Strategy: Support for research on the most effective methods for public engagement and education:
 - Evaluation and documentation of existing efforts at public engagement and experimentation with new approaches, with a particular focus on engagement with marginalized communities;
 - Community-engaged research to generate a diverse range of use-cases of how to leverage data to inform individual and community decision-making;
 - Exploration of approaches such as the European Union's "Citizens' Assemblies" on data as potential models for public engagement;
 - Exploration of impacts of broad public education campaigns to highlight data benefits such as New York City's "Brought to you by open data" subway ads;
 - Research into methods for protecting privacy for members of groups whose small population sizes may exclude them from data when small cell sizes are eliminated in some privacy methods such as differential privacy.
- Proposed Strategy: Support research to reduce disparities and increase equity:

There is also a need to understand how policies are implemented across different populations and whether those policies improve or exacerbate disparities. Without these key metrics, governments do not know how to increase benefits uptake for people in need, how uptake varies across populations or geographies, or how access to benefits impacts stability and health. This lack of data, combined with the challenges in state data infrastructure discussed below, limit the ability of policymakers, researchers, and the public to increase equity and improve outcomes for people in need of support. We recommend providing support for research and experimentation that will:

- Help to assess the ability of policies and programs to reduce disparities and to assess which programs, alone or in combination, are most effective at equitably increasing mobility;
- Help provide technical fixes to data bias that arises from imputation (consider NYU's Center for Responsible AI);
- Discover effective ways to increase coverage of undercounted or marginalized populations; and



- Assess federal data assets to identify gaps in data disaggregated by race, ethnicity, gender, disability, and other key demographics, with recommendations for different approaches to filling those gaps.
- Proposed Strategy: Increase transparency around data use and oversight of data-based decision-making tools.

Governments need support in creating transparency and disclosure about how they use the data they collect and determining when algorithms, predictive models, or other data-based decision tools are used to inform decision-making:

- Create guidance on ways to increase data equity for those who collect data, including state and local government, including permissible uses of federal data, best practices on transparency and disclosure;
- Support efforts to describe existing gaps or bias more accurately in data sets, with additional context on how the gaps or biases manifest in different use-cases;
- Support local communities in identifying new ways to use federal data to support local needs;
- Support efforts to append race, ethnicity, gender, and other demographic data to priority data sets; and
- Support efforts to expand algorithmic transparency, audits, and disclosures.

LEVERAGING NEW DATA AND CREATING NEW ECONOMIC MEASURES

A useful way to complement federal data efforts is to innovate and experiment with non-traditional data and new measures outside of government to identify any risks and benefits and to assess impact prior to considering potential full-scale government adoption. The experts noted two areas where the risks and benefits of new approaches are relatively unknown but promising: (1) leveraging private or other nongovernment data and (2) developing new economic measures to assess individual, family, and community well-being beyond traditional economic measures. Both policymakers and researchers spoke to the need to incorporate new data and to develop new measures of economic well-being beyond traditional indicators to gain increased insights into family and community needs, as well as understanding how programs interact with each other when individual family members receive different supports and how broader neighborhood or community factors may influence mobility.

Key Takeaways

• Broader measures of economic well-being could help advance mobility.

Economic well-being is traditionally measured through indicators such as the gross domestic product (GDP), unemployment rates, housing starts, consumer prices, and other indicators of inputs and outputs. While they are important components of economic activity, these traditional indicators do not capture broader indicators of stability and well-being that likely contribute to or impede upward mobility, such as health, segregation, poverty, social connectedness, civic engagement, and environmental and neighborhood factors. Creating a broader view of economic security and well-being, including both quantitative and qualitative data, would help address disparities and provide policymakers with valuable insights into the welfare of individuals, families, and communities. Researchers suggested a range of potential measures, such as including demographics like age, income, gender, race, and place, tracked across different agents in the economy, such as households, small businesses, communities, and more. In addition to traditional quantitative data, researchers pointed to the value of qualitative data, which provide more context and insights into daily activities and spending.



There is existing work underway in the European Union <u>exploring measures</u> such as work/life balance, education and skills, social connections, civic engagement, health status, and environmental quality that could be explored for possible translation to the U.S. Some experts similarly suggested leveraging the existing surveys, such as the Federal Reserve Survey on Consumer Finance to add broader questions to capture well-being, both topically and across different agents in the economy, such as family units, minority-owned businesses, and communities. Broader measures could help inform policy impacting which combinations of supports provide the best outcomes across different household make ups, the impacts of community and resources, and other factors. Ideally, some of the proposed experimentation can be done in active partnership with federal or other governments for their input and insights on the potential impact and pros and cons of different data or measures on the status quo.

• Non-traditional data offer new opportunities to provide more timely and granular insights.

The J.P. Morgan Chase Institute, Harvard's Opportunity Insights Economic Tracker, and a range of private sector data holders have been providing access to previously private data holdings in an unprecedented fashion to provide nearer-to-real-time data on the economic impacts of COVID-19. These private sector data releases have produced insights on the differing impacts of unemployment across income levels down to the neighborhood level—including how stimulus and enhanced unemployment benefits impacted savings rates and initially prevented an increase in poverty rates and how declines in housing payments could presage an eviction and foreclosure crisis, among other impacts. Additionally, many government services or programs are administered through nonprofit organizations or local governments, which could provide valuable information on who is accessing services such as workforce training or student coaching and how programs impact short- and long-term outcomes.

However, private sector data are not as nationally comprehensive as federal statistical data, and they are likely to have significant gaps in coverage of very low-income, unbanked, or marginalized populations. Because private sector data are primarily intended to inform business decision-making—or are commodified products in and of themselves—they are more likely to change over time in response to changing market needs, potentially undermining their usefulness for longitudinal research. Overreliance on private sector data supply can introduce additional risk to government data collection, if, for example, the companies that supply the data could go out of business or start to overcharge for their data once these data are incorporated into ongoing government analyses. Finally, many of the algorithms used to create private sector data sets are proprietary and not transparent, making it difficult to measure less visible biases in the data, and to determine whether changes measured using those data are real, or if they are a result of changes in the models being used by the data supplier.

Private sector data, although imperfect, could help address several of the unmet needs identified by policymakers, particularly in times of economic crisis—most notably timeliness and increased granularity across income levels and geographies. In times of economic crisis, timeliness is critical, and far too much government data lag by weeks or months, if not longer. Government data is also often aggregated at a state, regional, or national level that does not provide state and local governments with the information they need to target limited resources. In an effort to gather this kind of information quickly, the Census Bureau rapidly created and deployed the <u>Household Pulse Survey</u> in response to the COVID crisis. The Survey provides information on impacts such as food security, unemployment, housing instability, and other key measures, but it is aggregated at the state level and over only fifteen metro areas, limited its local usefulness.



Strategic Opportunities to Leverage Non-Traditional Data

• Proposed Strategy: External researchers and stakeholders should demonstrate new or high value uses of data to explore new economic measures, create new data "public goods", and inform federal processes.

The experts suggested a range of potential non-traditional uses of data, including leveraging nonprofit and private sector data, creating new metrics to assess and improve the delivery of government benefits and programs, and exploring different ways of measuring economic well-being beyond traditional indicators. Private sector data, particularly from financial data aggregators, could be used to track consumer spending, business revenue, and the flows of money in and out of bank accounts more frequently, and in a more geographically granular way than federal statistical data can achieve. Data from credit bureaus may also provide early indicators of financial distress, such as missed payments on credit cards, student loans, mortgages, or car payments. Indeed, the researchers noted that in some European countries, researchers can access credit data for free to explore these research questions.

Combining these data with federal or state data could help build broader measures of economic wellbeing, allow policymakers to provide assistance at earlier stages of financial distress to prevent worse outcomes, and broaden researchers' ability to assess core issues of stability and mobility. Exploring different ways of combining private sector data with federal data to maximize the unique attributes of each, especially in times of crisis, bears further exploration. Experimenting with these new approaches could help identify the risks, benefits, and potential impacts of different combinations of data to help focus the federal government on identifying whether, and how, to incorporate non-traditional data sources. Ideas included:

- Supporting research to identify what combinations of federal data with private sector data would best fill gaps in data frequency, timeliness, and granularity to answer key impact questions, while maintaining comprehensive population coverage, ensuring representation of undercounted groups, and clearly documenting gaps or potential bias;
- Supporting research to identify areas where new economic measures may provide better insights into economic well-being and to identify what data would be required to implement these measures, while clearly documenting what policymakers could expect to learn from expanded measures;
- Ongoing support for the creation of public data goods including private sector and government data combinations to allow for broader experimentation and documentation;
- Supporting the creation of additional public data goods including private sector and government data combinations to promote further experimentation and research; and
- Establishing partnerships or engagement with federal and potentially state agencies to help determine how—or whether—to incorporate private sector data into federal statistics and measure.

PART THREE: A STRATEGIC ROADMAP: ACCELERATING THE USE OF DATA TO ADVANCE EQUITY AND MOBILITY

At a high level, an ideal data system will include an effective feedback loop between data producers and data consumers—policymakers, researchers, program administrators, and the public—with data that is timely and geographically granular and which prioritizes the collection of data disaggregated by key demographics, reduces bias in data and use, ensures strong privacy protections, and undertakes frequent engagement with key stakeholders and the public to understand and respond to external demand. Advancing our current system towards this ideal will require an increase in internal capacity at all levels of government, improved technical infrastructure for the delivery of critical safety net benefits, increased data equity and public engagement, clear external demand signals, expanded data access, and support for diversified research, data innovation, and experimentation.

Based on the expert insights gathered over the course of the project, including the detailed suggestions on specific action items above, we propose the following strategic roadmap to accelerate the use of equitable data to increase economic mobility:

- 1. *Increase Government Capacity:* Increase internal capacity at all levels of government to collect, clean, link, analyze, and share data, using leading-edge privacy protections. Implement reforms to make government data easier and less burdensome to share and access, as well as more customercentric. Increase the capacity of governments to engage more effectively with the public, communities, and external stakeholders to inform data processes and supply.
- 2. *Increase Data Equity:* Increase data disaggregation of existing data and explore opportunities to link or append data sets to increase the availability of demographic information to measure and mitigate disparities; increase public and community education and engagement to increase awareness of data use and re-use, and prioritize community needs. Support research to help develop effective methodologies for public engagement and education.
- 3. *Expand State, Local and Tribal Government Collaboration and Data Sharing:* Increase collaboration across state and local governments to share data across regions and levels of government and provide a coordinated voice to engage with federal agencies. Work with local and tribal governments and communities to identify priority data, research questions, and data products to help meet local analysis needs.
- 4. *Increase transparency:* Increase transparency regarding data use, including expanding and accelerating release and discoverability of open, machine-readable data, public bias audits, disclosure, and oversight of the use of data-based decision-making tools in government decision-making.
- 5. *Create External Data Demand:* Support the development of strong external demand signals to inform government data improvement efforts across priority issue areas and from key stakeholders and to help ensure stable long-term funding and expertise for data infrastructure and analyses.
- 6. *Support External Data Experimentation:* Leverage external partners to experiment and innovate with high-value uses of new data, such as private sector and nonprofit data, combined with different government data to explore and document the benefits and trade-offs of different combinations in providing more timely and comprehensive insights. Support research into the development of new data measures to provide broader insights into individual, family, and community well-being and how these new measures may increase economic mobility.
- 7. *Expand and diversify research:* Expand funding support for research using government data, with expanded opportunities for a more diverse range of researchers and research questions.



8. *Strengthen the safety net benefits delivery infrastructure:* Support for efforts to increase the uptake of available benefits for key safety net services provided at the state level and for research into how state programs improve stability and mobility, individually and in combination.

Given the highly interrelated nature of the data ecosystem and the current gaps in effective data sharing and engagement, improvements in one part of the system are likely to have multiplicative benefits. For example, investments in increasing state capacity to share and analyze data, coordinate with each other across regions, and engage with local policymakers have the potential to improve the quality of data, data products, and responsive research to local governments, while increasing the likelihood that higher-quality data will be shared with federal agencies. Support for research on new data methodologies, new economic measures, or approaches to improve equity and engagement fill known gaps and are likely to be adopted and incorporated into governments increasingly aware of the challenges with data equity.

Improving the accessibility and quality of data has the potential to dramatically accelerate efforts to increase economic mobility and improve outcomes for populations who have too often been left behind. There are significant opportunities to increase data-informed policymaking at all levels of government, to answer core questions on which programs work most effectively for which populations, and to increase equity and public engagement. The increasing recognition of the need to improve data sharing across levels of government, to identify and address gaps and bias in existing data, to leverage data more actively from all sectors, and to bring communities that could significantly advance the cause of effective and equitable use of data to improve lives.