



A Locally Based Initiative to Support People and Communities by Transformative Use of Data

Julia Lane, David C. Kendrick, and David T. Ellwood
May 2018

The ideas in this paper were shaped by discussions within the Partnership but do not necessarily represent the views of all members.

The authors would like to thank the attendees of the [workshop](#) “A Road-Map to Transform the Secure and Accessible Use Of data for High Impact Program Management, Policy Development, and Scholarship” as well as the contributors to the resulting [volume](#) in the *Annals of the American Academy of Political and Social Science*: Beth Blauer, TC Burnett, Dennis Culhane, Peter Elias, John Fantuzzo, Ian Foster, Simson L. Garfinkel, Robert M. Goerge, Daniel Goroff, Matthew Hill, Jeffrey B. Liebman, Jeffrey Mays, Carla Medalia, Amy O’Hara, John Petrila, Jules Polonetsky, Nancy Potok, Andrew Reamer, Alexander S. Szalay, and Omer Tene. We are also grateful to participants in a follow-on [workshop](#) on the design of national data infrastructures, Robert Grossman, Mike Holland, Jeffrey B. Liebman, Amy O’Hara, and Andrew Reamer. Responsibility for any errors lies with the authors alone.

Responsibility for any errors lies with the authors alone.

ABOUT THE US PARTNERSHIP ON MOBILITY FROM POVERTY

With funding from the Bill & Melinda Gates Foundation, the Urban Institute is supporting the US Partnership on Mobility from Poverty. Led by chair David Ellwood and executive director Nisha Patel, the Partnership consists of 24 leading voices representing academia, practice, the faith community, philanthropy, and the private sector.

The Partnership’s definition of mobility has three core principles: economic success, power and autonomy, and being valued in community. Our collective aspiration is that all people achieve a reasonable standard of living with the dignity that comes from having power over their lives and being engaged in and valued by their community.

Contents

Executive Summary	iv
The Problem: Data That Could Empower People and Improve Programs Are Barely Being Used	1
Building Scalable and Sustainable Data Solutions	3
Lessons from Past Successes (and Failures)	3
Promising Places to Start: Building Research or Planning Capacity and Creating Real-Time, Integrated Service Tracking	10
Research and Planning Focus	10
Direct Services and Program Operations Focus	11
A Strategy for Moving Forward	13
Short-Term Demonstration Phase	13
Medium- to Long-Term Scaling and Sustainability Phase	15
Next Steps for Government, Philanthropy, Service Providers, and Others	17
Pathways to Prosperity: The Power of Data	19
Appendix. Successes That Can Be Built Upon	20
New Zealand’s Integrated Data Structure	20
Oklahoma’s MyHealth Access Network	21
The Coleridge Initiative’s Administrative Data Research Facility	23
Notes	25

Executive Summary

Effective use of data is central to any effort to increase mobility. The wise use of data amplifies and accelerates the impact of all other strategies. Scholars need data to understand the forces shaping mobility and poverty. Program managers need data to determine the effectiveness of different programs and to create opportunities for continuous improvement. And higher-level decisionmakers need data to inform policy choices, from which programs to expand or reshape to when and whether to target places or people for additional assistance.

Going beyond an operational level, data can enable innovation by making it possible to measure outcomes rather than just processes. As such, data can be the basis for embracing new programmatic ideas. Data can hold programs accountable for increasing mobility in its several dimensions. Data can help low-income people become empowered and families striving to make strong choices in their lives navigate the often-bewildering variety and complexity of “the system.” And, perhaps most important, data are a powerful tool for connecting the US Partnership on Mobility from Poverty’s other four reinforcing strategies for increasing mobility from poverty: changing the narrative, creating access to good jobs, ensuring zip code is not destiny, and providing support that empowers.¹ By using data to measure the success of people and programs, practitioners and scholars can understand the separate effects of very different efforts and begin to understand how they can amplify each other.

Fortunately, a great deal of potentially highly useful data already exists, much of it held by state and local governments. Unfortunately, much is held in ways that make it close to impossible to use effectively. From legal barriers to bureaucratic indifference or hostility, privacy concerns to stove-piped systems using widely divergent data structures, poorly maintained documentation to resource-starved technology, lack of training to perceived high political risks, the deck is stacked against those hoping to use data and information to transform how the nonprofit and public sectors provide services, similar to how data and information have revolutionized the private sector. There have been a great many efforts to build systems to use and link data effectively, but since they are often the result of painstaking and fragile one-on-one relationships formed to explore a specific issue, they are especially prone to dissolve when the parties or their interests change. This paper describes six core principles that we argue are essential for creating an effective and sustainable strategy to take advantage of the plethora of barely used data:

1. clear benefits to data owners;
2. empowerment and engagement of low-income adults and families;
3. security, privacy, and confidentiality by design;

4. standards for legal frameworks, technological capacities, and data structures;
5. strengthened human capacity, especially in state and local workforces; and
6. an up-front focus on sustainability, scalability, and a community of practice.

The paper then explores strategies for building on these principles. It offers two basic approaches: one with a research, planning, and policy development focus, and the other with an operations focus. It offers ideas for both short-term momentum building and a longer-term focus on creating a truly comprehensive and sustainable data system.

Impact on Three Dimensions of Mobility

The Partnership's definition of mobility has three core principles: economic success, power and autonomy, and being valued in community.

Investment: Unleashing the power of data requires financial resources for creating effective systems of privacy and access, training policymakers and key staff in both data building and effective use, improving metadata (e.g., documentation), building legal templates, and more. But the bigger investment may be in building trusted and enduring relationships across programmatic data holders in government and across insiders and outside scholars and practitioners. Some of this investment can and must be done by states and localities, but ultimately the effort will require investments in a national strategy to create standards and simplify mechanisms for appropriate data sharing.

Impact:

- **Economic Success:** An essential element of economic success is earnings, which can only be optimized through iterative program improvements, guided by data. All states routinely collect such information as part of their unemployment insurance system. Income data are also collected by federal and most state tax systems. Such data are at the heart of judging the success of programs ranging from training to support services like transportation and child care. School data can similarly indicate the success of early childhood and adolescent interventions.
- **Power and Autonomy:** Giving families a role in selecting which service to use while providing information on the impact of different alternatives would empower families. So would giving them access to information on their own service use. Even ensuring that identical information is not endlessly asked for in different programs would be an improvement.
- **Being Valued in Community:** Being given access to their own data and to information on what is available is a measure of respect and value. Overall engagement with and the building of community requires combining data from multiple sources ranging from health status to employment to schooling. Similarly, measuring the success of various program from training to criminal justice activities is essential for helping build community.

The Problem: Data That Could Empower People and Improve Programs Are Barely Being Used

The data revolution is transforming how people live, executives manage operations, and businesses deliver goods and services. Businesses use data to create better products, target receptive consumers, track performance, compare strategies, and improve efficiency—and in so doing create higher profits. Yet, when it comes to helping people escape from poverty, the revolution has barely begun. Several idea papers from the US Partnership on Mobility from Poverty refer to the need for more or better data to maximize the potential of new proposals.

The value can be enormous. Just as businesses use data to create better products, government agencies could use data to design better programs at lower cost. A project that links data on training programs to unemployment insurance wage records could help determine what works and what doesn't in terms of earnings and employment outcomes. Just as the private sector finds the right consumers for its products, the public and social sectors could serve people better by building a real-time, fully integrated, secure system that allows providers and families to engage effectively with the whole person and whole system instead of the stove-piped, program-by-program, case-by-case system found nearly everywhere in government and communities. A caseworker reviewing a foster child's files with his or her family could discuss the specific health issues that might be interfering with the child's progress at school and connect the family with resources in the community.

The raw data to do this exist. There are vast amounts of administrative data in federal, state, and local governments and communities throughout the country. Philanthropic foundations and federal agencies have spent millions of dollars to attempt to draw on such data, yet there are too few success stories. The reasons are well understood. There are fundamental structural problems: the startup costs to access and use the data for policy purposes are dauntingly high,² and the rewards to individuals are too low. In prosaic terms, the data plumbing needs to be installed before the evidence house is built, and investment in plumbing has been lacking. The result has too often been a series of one-off and artisanal approaches, rather than a serious, coordinated investment strategy. This paper draws on the experience of many researchers and practitioners to explain why and to suggest what foundational investments need to be made to avoid the mistakes of the past.

There are many reasons for the often ineffectual use of government and private data. One is the variety of sources of data. People in poverty interact with largely independent program and service delivery

structures, each collecting and defending its data in its own idiosyncratic ways. Each program typically has its own rules, collects and maintains its data separately, deals with its own clients, and targets different outcomes. Often services and programs within the same system have unique and separate data. For example, it is extraordinarily challenging to piece together a person's criminal justice history. As Mueller-Smith has found, there are thousands of data sources in state-level criminal justice agencies, including police, administrative offices of courts, departments of public safety, and criminal justice commissions, with no standardization or unique identifiers across systems.³ In practical terms, this means that a first-order investment is to develop tools to integrate and standardize data to make it useable across agency lines.

It has been said that interoperability happens at the speed of trust. Unfortunately, there are strong disincentives to share data across programs or with outsiders, and strong governance built on relationships and transparency is required for building the necessary trust. Indeed, there are strong perceived economic, legal, bureaucratic, and political risks to sharing data. The stewards of public data, including leaders of private and nongovernmental organizations that receive considerable public financial support to provide services, perceive high risk for loss of privacy, legal liability, adverse media, political upset, damaging or unfair revelations, and external meddling. Private providers (such as health care practitioners and vendors) perceive data sharing as expensive and potentially threatening to their businesses. People whose data would be tapped rightly worry that they will be tracked and categorized in ways that reduce their autonomy and stigmatize them.⁴ And even if the parties wanted to share information, current policies and the available resources, technology, and human capital are often wholly inadequate to the task. Addressing these barriers at scale is critical to success.

There needs to be clear value generated for data providers.⁵ Such a focus should be paramount, so each barrier can be addressed and the political will to do so can be generated.

There is, however, reason for optimism. Other domains have successfully built data infrastructures, and their experience can help inform the task at hand.⁶ Within the federal system, such successes have been achieved with great effort and financial savings at the Census Bureau and the Internal Revenue Service, and more may be on the way.⁷ State and local governments are starting to work with each other, with universities, and with think tanks to link education, human services, criminal justice, and workforce data. Comprehensive health data from multiple providers is now being linked in real time in many parts of the country.⁸ If we are to make use of the data required to understand poverty and enhance mobility collected by local and state governments—including data on income, employment, health services, crime and incarceration, program usage, education inputs and outcomes, job vacancies and hiring, unemployment, housing, and child support—it will take a similar investment of time and energy. The same issues will hold with using data held in the private sector, such as health care, social services, and organizations funded by private and religious philanthropies.

Building Scalable and Sustainable Data Solutions

Many groups are beginning to see the sizable possibilities for using existing data to enhance mobility and well-being. There are dozens of compelling examples of places where selective data from various administrative sources have been put to effective use. Too often, however, these examples are unique, one-off arrangements between a stakeholder (such as a program leader, policymaker, or philanthropy) and researchers to investigate a single issue. Each such effort crafts unique strategies to solve issues of trust, legality, privacy, technology, governance, and financing—in that one case. The problem with “artisanal” data uses is that they are almost impossible to reproduce easily and sustain financially. By drawing the lessons from these existing efforts, future projects could help make data far more useful and durable for practitioners and scholars and empowering for families. And we need a strategy that moves us from a wide variety of differentially successful programs to something that will dramatically accelerate progress in this domain.

Over the past 18 months, we have sought to learn about the many efforts and convened two workshops that identified existing successes in the social and natural sciences and distilled the essential elements of those successes. The lessons have helped identify the power and the practical elements of scalable and sustainable systems. It should be possible to stimulate new systems and create the economic, technological, and human foundations for rapid expansion and deployment across the country.

Lessons from Past Successes (and Failures)

In several demonstrated cases, scalable and sustainable strategies have emerged. Here are a few that we know best.

The Longitudinal Employer-Household Dynamics (LEHD) project linked unemployment insurance wage records for 49 states. These data were also linked to data from the Census Bureau, the Social Security Administration, and the Internal Revenue Service. The result is a powerful dataset that provides information on the dynamic interaction of workers and firms. It now serves as a vital tool to benchmark employment and earnings as a basis for development strategies, for planning transportation infrastructures, and more. It has been used to trace the effect of training programs, immigrant entrepreneurship, earnings inequality, and a host of other issues. Indeed, the microdata now represent one of the most heavily used labor market datasets for researchers.

The program did not, however, miraculously appear. It was made possible because of the operational commitment of the Census Bureau, the vision of a few dedicated people, the engagement of key

stakeholders in state government, the sustained funding of foundations (notably the Alfred P. Sloan and the National Science Foundations), and the development of products that have great value to stakeholders. The program began as an idea in 1997; it reached full fruition some 15 years later and is still evolving.

In Oklahoma, MyHealth Access Network began life in the health care sector as a health information exchange, helping more than 5,000 physicians and nearly 100 hospitals deliver higher-quality, better coordinated, and cost-efficient care to more than 3.5 million people. In 2012, MyHealth convened the Oklahoma instance of a Medicare demonstration project called the Comprehensive Primary Care Initiative; over four years, the program saved more than \$100 million in the care of approximately 50,000 patients.⁹ The foundational trust, governance, and technology created by MyHealth is now being embraced by behavioral and mental health, social services, and even corrections and education programs to tackle Oklahoma's problems in new ways. In addition, MyHealth has connected with other health information networks around the country to create patient centered data homes, which currently cover 50 million Americans and is expanding to include as many as 230 million within the next two years.

In several jurisdictions, states are sharing multiagency data on formerly incarcerated people and people receiving public benefits linked to earnings data in a cloud environment. Just like LEHD, the infrastructure was made possible by a major commitment from the Census Bureau, the Laura and John Arnold Foundation, and the Overdeck Family Foundation; the hard work of a few visionary individuals; and the engagement of state and local stakeholders. An important additional feature has been the incorporation of training programs: the data can be shared, and staff in contributing agencies are empowered and trained to use and access the data.

New Zealand has built an integrated data infrastructure that enables early interventions in the lives of disadvantaged individuals. As one New Zealand service provider notes: "The more information you have the better choices you can make. It's not just the dataset ... but it's that dataset in conjunction with other data that we've got—bus routes, schools, where we're getting kids from. It gives us insight and the opportunity to tailor and make choices around how we deploy our services and how we engage with other providers."¹⁰ In 2001, a determined group of people within New Zealand's government became aware of the developments of the LEHD program in the US. They went door to door to government agencies charged with providing social services, and convinced them of the value. An initial investment in linked employer-employee data by Statistics New Zealand evolved over time to demonstrate that value could be realized, and resulted in the program as it exists in 2018.

We sought to distill key lessons from programs that have been tried across the country and around the world.¹¹ We conclude that six principles should lie at the heart of any major data initiative for mobility from poverty:

1. *Clear benefits to data owners:* Data producers and owners must see significant value in sharing access to the datasets. Data owners must build trust to ensure their mutual benefit and even-handed participation.¹²
2. *Empowerment and engagement of low-income adults and families:* Some of the data that can help vulnerable populations the most can, in the wrong hands, also harm them. This possibility must be prevented and the benefits of the data maximized for real people, not just research studies. One strength of private-sector data use has been that customers see it as benefiting them with customized, targeted access to the goods and services they want. And increasingly, companies are giving consumers greater control in limiting who can see their data. The data must be usable to help empower low-income families and substantively improve service delivery at the community level.
3. *Security, privacy, and confidentiality by design:* Both technical and policy aspects of privacy and confidentiality must be included by design, not developed ad hoc or as an afterthought. State-of-the-art systems and protocols are essential.¹³
4. *Standards for legal frameworks, technological capacities, and data structures:* Scalable systems require standardized templates for legal language, data models, syntax and terminologies, and domain-independent identity resolution standards must be selected and employed to support an architecture that enables local, regional, and potentially national data aggregation.¹⁴
5. *Strengthened human capacity, especially at the state and local levels:* Human capacity at local or regional levels must be strengthened to support (1) the development of governance and policy, (2) the technical skills required to assemble and organize large multidomain datasets, and (3) the analytical skills necessary to explore the data for new insights leading to new interventions and monitor the progress of new programs.¹⁵
6. *Up-front focus on sustainability, scalability, and a community of practice:* Sustainability and scalability must be central foci, in part by creating communities of practice that develop nationally credible and practical best-practice standards. Any demonstrations must offer a clear plan for expansion and economic sustainability.¹⁶

We briefly consider each principle below.

Clear Benefits to Data Owners

One of the most serious challenges to making more intelligent use of government and private-sector data is that the data are typically produced and then held by programs and organizations designed for a single

purpose. Even when “moving individuals and families out of poverty” is the mission of the agency or organization, state policy, privacy and confidentiality concerns, and resource constraints are frequent barriers. Ideally these data owners would recognize that they could advance the objectives of their programs far more successfully by creating managerial tools, mechanisms for determining what is working and what is not, and even testing innovations. That requires a demonstration of value and an authorizing environment committed to more effective use of information.

Equally critical is building trust across agencies and data owners, as well as with potential users of the data. If owners believe their data will be put to mutually beneficial uses and that their programs or clients will not be compromised, they are likely to be open and even enthusiastic participants. But if they do not trust the participants and the uses, they can delay and often block almost any effort. Building trust requires effectively convening stakeholders across communities and regions to establish governance for data exchange and use, as well as the ongoing leadership to maintain the delicate balance of trust required to ensure the continued flow and productive use of data.

Empowerment and Engagement of Low-Income Adults and Families

Much of the data of interest relates directly to serving low-income families. Unfortunately, fragmentation of the agencies and systems intended to deliver services to these families also produces fragmented, low-quality data. Thus, the quality, comprehensiveness, and convenience of services are in question, and our ability to study and thus improve them is constrained. But many advocates fear that integrated data will be used to further regulate, isolate, and stigmatize low-income people. The benefits and the risks of targeting specific individuals are lower when deidentified data are used. When using client-facing, real-time data for identified individuals, both benefits and risks are amplified.

The essential goal is to ensure that such data empower and strengthen low-income families and individuals. Ultimately, as is becoming common in health data, program data should be securely available to families so they can better navigate and effectively use available services and correct errors and abuses. With permission from families, coordination services like community resource navigators should be able to monitor the entire process and offer support and direction, or even facilitate the next connection to services. Similarly, downstream providers or services could be granted access to improve families’ experiences and tailor services to their needs. As families move from one service provider to another, being asked to provide the same information over and over is enormously burdensome and invasive. This information could and should be gathered once and used immediately to deliver improved services. Finally, these data, which are a by-product of the actual work, can now be used to study and improve the services at a system level.

Security, Privacy, and Confidentiality by Design

Security, privacy, and confidentiality are critical concerns for all stakeholders to data systems. Families and their advocates fear that data could be used to adversely target and stigmatize them. Data owners may fear the liability risk presented by sharing data, and data users may be confused about approved data uses.

Answering these concerns requires a combination of governance/policy achievements and technical security measures. Necessary governance and policy achievements include establishing a transparent, representational governance structure to promulgate policies that require documentation of client consent and clearly outline acceptable data uses. These policies should accommodate the relevant provisions of HIPAA, FERPA, 42 CFR Part 2, and other relevant privacy laws at the state and local level.¹⁷

Fortunately, technology has made enormous progress with measures to ensure that data are kept as secure as possible and that their use is safe at several levels. Investments, baked in from the start, in a combination of state-of-the-art technical strategies and thoughtful human oversight and screening could dramatically improve privacy and usage protections. Various standardized mechanisms have been developed for different confidentiality situations, ranging from deidentification, to mathematical mechanisms to add noise to data, to secure enclaves (ultimately in the cloud) with mechanisms for certifying safe users, safe analyses, and safe products. Legal hurdles, often written when current protection methods did not exist, can be removed while increasing security by developing more up-to-date templates that could be widely used.

Standards for Legal Frameworks, Technological Capacities, and Data Structures

Best practices from many industries indicate that data structural and content standards are critical to effective linking and use. Legal barriers are often cited as primary reasons that data cannot be shared. Standardized legal templates for data access and use are essential for progress. States and localities often are reluctant to let others store their data for security reasons. In fact, all the technology experts at a roadmap conference convened by this project argued that cloud-based enclaves often provide the best and most cost-effective protection for data storage and ease of access. Administrative data often lack clear documentation of the content, quality, and use of the data fields. Often there is not even a data dictionary. Computer scientists can build ways to capture routinized and reliable metadata, as well as ways for users to offer feedback on quality and problems—just as Amazon and TripAdvisor have done for large-scale private-sector data.

Finally, and vitally, demographic data elements need to be standardized and analytical methods applied to uniquely resolve individual identities across multiple data sources. Successful linking of the data requires

that a minimum number of demographic variables are available and that they are correct in content and format. Once correctly linked, the data can be deidentified for inclusion in larger research efforts. Only by adhering to data standards and accurately resolving identities will the validity and dependability of conclusions drawn from the data be preserved.

Strengthened Human Capacity, Especially at the State and Local Levels

Investments need to be made in people as well. Even with the right systems, the agency workforce needs technical expertise, and fluency in data science and basic analytical tools in agencies is extremely limited. Agency leaders need to see concrete evidence that linking data across agency lines will advance their agency mission and the people they serve. This leads to a recurring challenge: data must be assembled before proof of value can be demonstrated, yet leaders often hesitate to share data before the value is clear. This challenge has been addressed with data testing agreements that limit data use to the proof-of-concept case; having sophisticated data experts in each organization greatly speeds the process.

Building workforce capacity in the relevant areas of analytics, privacy law, confidentiality protection, and computer science forms the bedrock for initiating data aggregation and for sustainable success. State-of-the-art executive education programs are emerging now. Standardized curricula could be developed in each of these areas.

Up-Front Focus on Sustainability, Scalability, and a Community of Practice

The financial, political, and organizational sustainability of each community's data efforts is critical to consider very early in the process. Many demonstrations predicated on the assumption that "if we build it, they will come *and pay*" have not built economically sustainable strategies. Many programs funded by foundations or by scraping together short-term budget authority flounder without a credible long-term financing plan. Early wins to establish the benefits and credibility of data systems do require up-front funding. But planning should also require a credible system of ongoing long-term financing. That will often require navigating the complicated politics of government funding, but data costs are small compared with the costs of the actual programs. And, ultimately, a routinized and standardized system used across programs could save considerable money not only in service costs, but also in the data collection and processing costs.

A final and particularly vexing challenge is replicating the results of a successful program in other communities and populations. Scalability of the program beyond the community should also be considered early. One-off technology and architecture strategies can become trapped in their uniqueness without clear

pathways to a larger and institutionalized system. This prevents the program from achieving the added value of new data sources and loses the economies of scale afforded by leveraged technology, expertise, and trust agreements. It is far cheaper to add to a strong existing system with clear standards and policies for data than to constantly reinvent new strategies.

Thus, while programs are being designed for some immediate successes, they must also be part of building communities of practice with the goal of establishing trusted collaborations between program operators, researchers, and stakeholders. Such communities can also be used to set and maintain high standards for data quality and metadata, credible ideas for innovation, development of new tools, and ensuring the reliability of research and evaluation results.

Promising Places to Start: Building Research or Planning Capacity and Creating Real-Time, Integrated Service Tracking

It is helpful to think about and learn from two successful approaches to building community data capacity. In each case, the goal is to build integrated data systems that provide actionable data. In one case, the focus is on research, evaluation, and policy planning. In the other, the focus is to provide direct client-facing services. This section describes how a research/planning focus or an operations/direct service focus can be put in place.

Research and Planning Focus

The research and planning end of the spectrum is exemplified by the LEHD program and the Coleridge Initiative, as described in the appendix. This strategy recognizes existing data are invaluable and that one can build trust, capacity, and momentum in learning and using that data to find answers to specific questions.

Some of the loudest calls for better data use have come from researchers. The work of Raj Chetty and his colleagues is an excellent example: they were able to use IRS data linked over many years to explore the extent of economic mobility across generations and to determine the role of place and other factors in influencing it. Their recent success in matching the data with Census information has made the insights even more compelling.¹⁸ Researchers also often use administrative data to evaluate programs (especially innovative ones) by tracking their impact on outcomes ranging from employment, to children's school achievements, to health, to arrests. These outcomes can be found in existing administrative records, such as employment information from unemployment insurance and school performance from school records.

The Moving to Opportunity experiments, which offered low-income families the opportunity to move to higher-income neighborhoods, sought outcomes along all these dimensions.¹⁹ Note that data linked in this way are not designed to track and guide immediate service decisions for individual families. Rather, they are used primarily to understand larger structural patterns or evaluate program impacts. Such data do not have

to be available in real time. And once the requested data have been linked, researchers can create and use a stripped, fully deidentified dataset that protects the privacy of the individuals.

More recently, many governmental units have begun seeing the potential for developing powerful data analytic tools that can help answer important policy design questions and inform individuals about the effects of different choices. Having a functional way to access existing administrative data can open a large range of big-data techniques for understanding the incidence, costs, and potential impacts of new policies. Those techniques become ever more powerful as new datasets are linked in. Research- and evaluation-driven strategies and program/service strategies require that high-quality data be maintained, linked, stored and protected, and shared under appropriately protected conditions. But the former typically does not require data in real time; indeed, it often takes months or years to clean and match the data. Moreover, time can be taken to screen the researchers, their proposed uses of the data, and protect the results they hope to publish.

There are many examples of the value that can be generated as a result of such approaches. Transportation planners and Workforce Investment Boards use the On-the-Map and quarterly workforce indicators generated by the LEHD program to plan transportation routes and support programs for low-income people. States are developing apps to show students the earnings and employment outcomes of community college specializations.²⁰ As Liebman argues, research can help target resources—for example, to improve outcomes for the entire target population, rather than focusing on those who arrive at the agency’s front door. And data and training can help agencies answer broad programmatic questions such as “are large caseloads all from a single regional office that is understaffed? Is the heterogeneity in caseloads across caseworkers appropriate with some caseworkers given a larger number of low-intensity cases and others given a smaller number of higher-intensity cases? Is it resulting from some caseworkers failing to do the paperwork to close inactive cases promptly?” and so on.²¹

Direct Services and Program Operations Focus

At the other end of the spectrum is data development driven by the needs of program and service operators. Of the programs listed in this paper’s appendix, MyHealth is the clearest example. Such an approach allows service providers to serve their clientele more effectively through access to data about their use of various services and how such use affects clients’ employment, schooling, health status, and/or involvement with the criminal justice system. When offering direct services to individuals and families, data need to be local, timely, and actionable. If a doctor is serving a patient, he or she needs to know the previous test results and past services while the patient is being treated. The doctor can also use the captured data to explore which treatments prove most effective. Similarly, “coaches,” navigators, or caseworkers would have a much clearer

picture of a person's history and progress if they could review current and past employment, health crises, and many other indicators.

Such data might be used to measure the success of whole programs or individual providers and caseworkers. In many ways, the availability of real-time individual data closely corresponds to the data revolution in private industry, which increasingly involves using timely information to target products, monitor quality, provide customer service, and measure the performance of everything from individual products to sales and expenses for individual sellers to whole divisions.

In such settings, the identity of the person whose data have been linked must be correctly resolved across multiple sources and available only to appropriately credentialed providers. Note also this latter system can also allow the people being served to see their own data in real time—that is, see their medical record, test results, and appointments or, in a private-sector setting, recent purchases, services, and pending issues.

The challenges of creating such systems are considerable. The need for near-real-time treatment and outcome data requires integrated and linked data systems in which high-quality data are entered quickly and accurately and where users can be given trusted status (by the program and/or recipient) to allow them to have access to someone's identity. Such information has enormous potential for vastly more integrated and immediate help. But it also raises fears that individuals will be targeted in ways that might invade their privacy and inhibit their legitimate use of services.

The building of real-time, client-facing data systems requires a very high level of trust and ongoing commitment, particularly when asking different programs and providers to share data on the same clients. Governance and leadership are especially critical since the challenge of building effective systems is great, and it is politically and operationally impossible to put one program or provider in control of the entire system.

In principle, these two different focuses could ultimately arrive at a common unified system, where detailed identifiable data are potentially available in near-real time for those who have been granted access while researchers, overseers, and senior leadership can be granted access to aggregated or deidentified data for appropriately screened purposes. That seems to be the case for New Zealand. But it took decades to get there, and New Zealand is far smaller and more homogeneous than the US.

A Strategy for Moving Forward

Our review has made clear the reason for the lack of data to support the goals of the Partnership. A sustained focus on building a data infrastructure—the plumbing—is necessary. So is the building of value and trust. That involves both substantive investments and an operational design for a scalable national infrastructure.²² That infrastructure should be designed to encourage communities to invest in valuable ideas and should be substantial enough to promote the scaling of successful approaches across state and local boundaries. Building value and trust also involves designing systems of governance and engagement that lead to excitement, rather than fear, of what the future can hold.

We believe engaging a group of willing and committed communities can generate highly visible reforms and successes. The goal should be to establish local governances that leverage proven policy templates and data standards, as well as effective privacy and access protections. The value of such a core infrastructure would be championed by credible “data shepherds” who would be identified and charged with helping build out to additional communities and data sources. There would be a short-term demonstration and learning phase, followed by a longer-term sustainability, scaling, and community-building phase.

Short-Term Demonstration Phase

Rather than building yet-another new effort, a far more promising strategy is to **build on emerging models with real promise**. Most communities have champions for more effective use of data who have achieved some success. One such example is the MyHealth Access Network referenced earlier and described in more detail in the appendix. MyHealth has created real-time comprehensive health records that serve most Oklahoma consumers and providers through the provision of actionable data to thousands of clinical and social services encounters a day. The network has a trusted system of governance with government and the private-sector stakeholders deeply engaged as well as a close link to consumers and employers. Both the network and governments have begun linking social services, public health, early childhood, and other government administrative sources. And MyHealth has a highly functional set of analytic tools. The challenge and opportunity in this case would be to expand the data sources sustainably into high-priority domains for upward mobility from poverty including employment and earnings, human services, and potentially education and other privately provided services. Moreover, as it has done with other projects, MyHealth’s governance would expand its engagement with low-income families to ensure their voices guide the program.

The research-focused approach is embodied in the New York University Administrative Data Research Facility (ADRF) also described in the appendix. This program was built around the Census Bureau's very successful LEHD effort, which enables multiple local, state, and federal agencies to keep data in separate secure areas, approved by FedRAMP protocols, and link data for approved projects. The ADRF allows agencies within the same state or different states to agree to share their data in a common area in the cloud for specific approved projects. In this case, administrative units can put their data in a protected domain. When mutually agreed, data can be easily linked and analyzed. Confidentiality and access protocols are well developed. One particularly important element of this strategy is that it invests heavily in the training of state and local staff in the skills and potential of the data systems. Local consortia can evolve differently in the ways that they govern and collaborate with the data.

After building on a promising model, the demonstration phase would focus on three goals:

1. **Create highly visible successes.** Credible examples of quality implementation of linked data are critical to success. National leaders would build regional data infrastructures around health, employment, and low-income populations. They would be the initial visible faces of the initiative, charged with providing poster examples of the value of linked data and offering best practices and successful techniques to house data and provide access. A critical element would be developing a training curriculum to engage federal, state and local agency staff in the secure linked environment to build common documentation, measures, and products across agency and state lines.
2. **Shepherd local projects.** National leaders would be identified and be responsible for recruiting, fostering and growing the initial projects, and with providing those projects with access pathways to data providers, codifying lessons, and beginning to build a community of practice that will be essential to accelerating efforts across the nation by examining voluntary standards, data and legal templates, and more. They should be charged with providing examples of proven and developing governance models, policies, and technologies for data access and use.
3. **Establish ongoing evaluation and learning.** Using both project leaders and outside groups, including independent advisors, the project could determine the impact of the initial projects on mobility. It would also examine the level of engagement, and sources of excitement and concern among "data owners," low-income families, advocates, local government, business, researchers, and other stakeholders. It could review the success of alternative confidentiality arrangements. It would focus on the most effective strategies for producing the human capital necessary for the effort and the most powerful technological and analytic tools.

Medium- to Long-Term Scaling and Sustainability Phase

In a second phase, an expanded initiative could conduct a highly visible multistage competition for the most compelling local upward mobility programs that use evidence from multiple domains. The innovations would be required establish data governance, develop policies and use data to track outcomes at multiple levels, show evidence of value, and be replicable and scalable.

The initiative would **provide ongoing support** beginning with initial proposal stage, through the competition, and to the winners as they implement their ideas. There would be three elements to that support:

- **Access.** Positive outcomes will not be possible if data cannot be used effectively and protected. The initiative will require that all projects build or participate in state-of-the-art interoperable data infrastructures. This may include the creation of regional data hubs using private vendors to provide low-cost, high-quality data storage and protection.
- **Training.** Capacity must be developed at all levels if innovative approaches are to be sustained and expanded. The initiative will both develop and support capacity-building programs as well as develop ways to encourage community tools to document data.
- **Leveraging.** The initiative will seek existing government and philanthropic programs to leverage at the program level and encourage applicants to do the same by weighting leveraged funds and resources into the evaluation.

The expanded initiative would also **create national sustainability**. To avoid the hazards of past one-off “artisanal” efforts that yielded positive results but failed to stimulate others to follow suit, the impact strategy would include the following:

- Focus on projects that are likely to be highly visible with clear wins for the data “owners,” both public and private, as well as governmental leaders.
- Build a learning community of stakeholders. That would include working with groups at the community level who have established effective multi-organizational governance, including the National Governors Association, National League of Cities, as well as governmental chief information officers, technologists, scholars, and vendors. The group would also include the winners and the promising but unfunded groups from the competition.
- Provide a strong voice of information and advocacy. The initiative with its many elements will work to ensure that successes are amplified and that threats to creating effective data systems are disarmed.

- Offer and evolve project standards to link successful projects to one another so data assets and use cases may be exchanged among communities and across domains of service to maximize their impact.

As with demonstration project, the expanded initiative will support **ongoing learning and evaluation**, with a particular focus on the impacts and the achievement of the six sustainability principles noted on page 5: real benefits for data owners, empowering low-income families, privacy by design, building human and technical capacity, and ensuring scalability and sustainability).

Ultimately success will be measured by whether data systems become an intimate and essential part of helping far more people in the US move out of poverty.

Next Steps for Government, Philanthropy, Service Providers, and Others

Governments hold much of the existing data and establish many of the data use policies for both public- and private-sector data. Government leaders and the providers of government-supported services who use data across different programs and agencies can vastly improve how they serve low-income individuals and families, improving their lives and potentially cutting costs to taxpayers. And data sharing could be the start of greater collaboration, enabling treatment of people as individuals with unique experiences and needs across multiple domains, rather than simply focusing on a program in a single domain serving the masses, disconnected from and uncoordinated with other programs.

Some government agencies are anxious to achieve these goals. They are investing in some key elements—their technology and their workforce—necessary to effect change. They are reaching out to low-income adults and families and their advocates to ensure they are empowered rather than overregulated. And they are reaching out to scholars and practitioners to help ensure that the power and quality of the data are high and that data can be shared effectively with those working to increase mobility. Still, too often these efforts have been localized, artisanal, and one-off.

Other governments can and should follow the lead of these early adopters and learn the lessons (both positive and negative) from these ongoing efforts. Governments need to commit to a long-term, more comprehensive, and ultimately sustainable structure for properly collecting and taking full advantage of their program and other data. Yet the ability of agencies to do so is limited for several reasons. There are many legal, technical, human, and financial hurdles to clear and few prototype successes to point to. The pressures to meet existing program needs make it difficult for agencies to try something new and create pipelines of new products. Government salary structures impede the ability to hire and retain enough in-house data analysts, so agencies don't have the capacity to work with new linked data. These combined challenges have led to the current situation: agencies cannot get the significant resources necessary to make use of new data, and because they don't use new data, they don't get new resources.

The role of philanthropy can be particularly critical in breaking this negative spiral. To succeed, philanthropy must encourage long-term generalized and sustainable data solutions, rather than the more common use of external funding focused on getting one-time data on a particular issue. Philanthropic funding capital can provide flexible seed funding to spur initial collaboration. Local and national philanthropies can serve as neutral third-party conveners helping to pull together political and community

leaders, key agency heads, providers, and advocates to build mutual trust in the design of more powerful systems at the state and local levels.

Ultimately, philanthropy, particularly larger national funders, can also play a vital role in lowering the cost of building effective data systems. They can help create standards and mechanisms that simplify legal templates, routinize strategies for privacy and data protection, and build common elements for key variables. They can provide initial resources to support and/or build secure state-of-the-art technological systems that permit agency staff and researchers to safely access and link public and private data, and to analyze data across agency lines; and (ii) build staff workforce capacity to work with the linked data. The role of cross-government and cross-agency training programs and other trust-building activities is critical; they build a network of invested agency staff who understand the potential power of data, including strategies for creating quick-win and long-term products, along with the recognition of the hard work required for ensuring accuracy and fitness of the underlying data system.

Universities can also be a crucial player. Just as universities have served as reliable sources of information as part of the agriculture extension service, so too can universities offer state-of-the-art expertise and training on neutral ground. Universities offer a potential place to bring together parties with different interests and to shepherd data projects forward. Many of the most effective activities using state and local data—from criminal justice work to human services to employment and earnings—are university based. Scholars can help ensure that the data are properly cleaned and integrated and that high standards are maintained in the analysis. In some cases, university consortia can serve as holding areas for data and facilitate appropriate safeguards and rules of use. Scholars also can serve as useful advocates for the power and limitations of information and help set priority areas for improvement.

Finally, outside service providers, community organizations, private-sector leaders, and other institutions also need to be involved in demonstrating value. Providers can indicate which data are most needed and ensure a workable system of use is devised. Community organizations can ensure that low-income residents are being protected and empowered, and they can help indicate places where early wins might be achieved. The private sector is much further along in using data, understanding state-of-the-art solutions, and recognizing the need to build long-term sustainable solutions. This progress should be studied and the best practices should be enlisted in the effort to increase mobility from poverty.

Pathways to Prosperity: The Power of Data

Data are an essential part of creating genuine mobility from poverty. The success of the US Partnership on Mobility from Poverty's strategies depends on quality data. Data are not only vital for measuring success and failure. They hold potential for providers to learn and innovate and improve. And, perhaps most important, data provide the key to giving struggling families and individuals real power to determine the most promising pathways for their efforts to move out of poverty. Our strategy is designed to build on leading examples of success and create rapidly scalable and sustainable data systems with the capacity to transform lives.

Appendix. Successes That Can Be Built Upon

New Zealand's Integrated Data Structure

The vision has become real in other countries, but it takes time and persistence. The Integrated Data Infrastructure²³ in New Zealand was started in 2001.²⁴ Figure A.1 provides an overview of the structure in 2017.

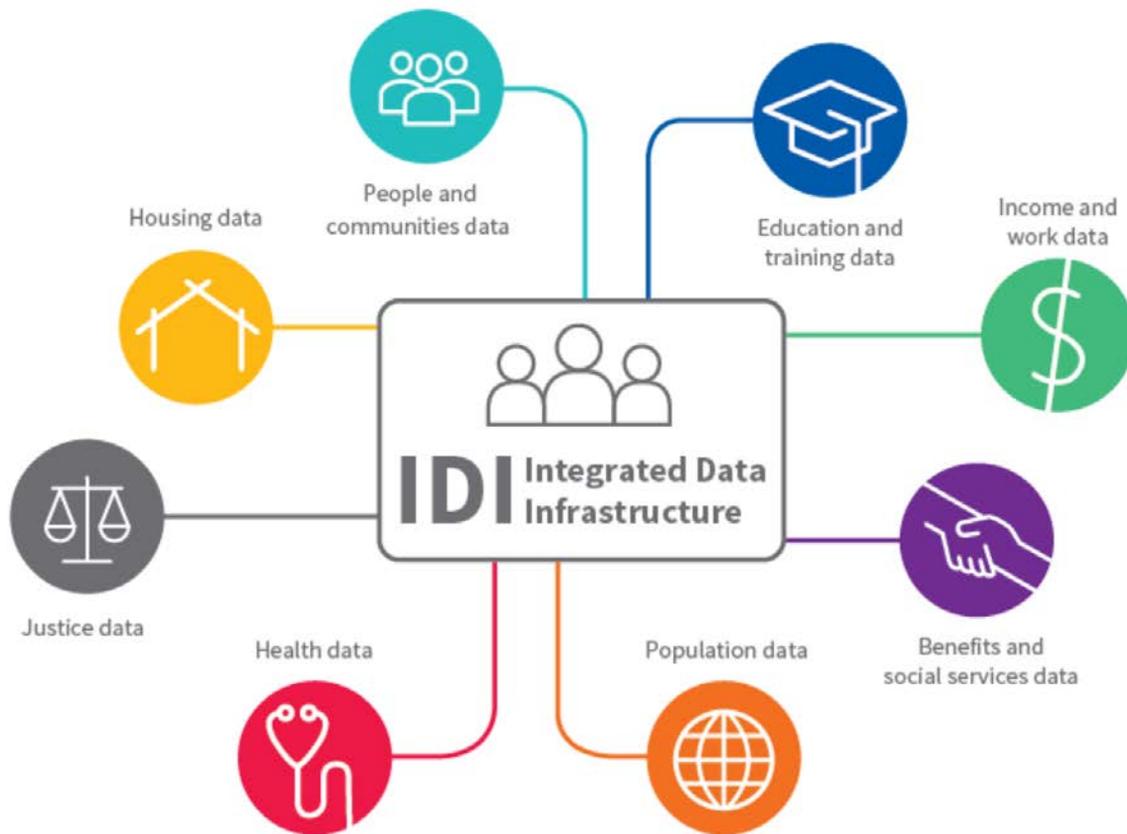
Examples of the way in which income and work data have been used to help improve New Zealanders' access to "good" jobs include these two:

- Careers NZ's Compare Study Options (<https://www.careers.govt.nz/tools/compare-study-options/>) helps young people make better decisions about where their study choices can lead them. The Ministry of Education created this tool by using combined student loan, tax, and education data.
- Training providers use the research to improve employment outcomes for youth, such as detailed information about where to hold job fairs and what kind of training to provide.²⁵

Probably the best indicator of the success of the program is that the approach is now being emulated in Australia with a \$A35 million investment by the federal government.²⁶

FIGURE A.1

The New Zealand Integrated Data Infrastructure



Source: Stats NZ. Licensed by Stats NZ for reuse under the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license.

Oklahoma's MyHealth Access Network

MyHealth Access Network is an Oklahoma nonprofit that began life in the health care sector as a health information exchange, serving more than 5,000 physicians and nearly 100 hospitals to help them deliver higher-quality, better-coordinated, and cost-efficient care to more than 3.5 million people.²⁷ Over time, other sectors, including public health, behavioral health, disability services, social services, first responders, and now early childhood education, have joined the MyHealth effort, enabling the assembly of a unique dataset. By providing a robust infrastructure for convening and governance, as well as practical data exchange and analytics services, MyHealth has helped Oklahoma become the site of Medicare's most successful innovation demonstration project, reliably achieving 5–7 percent annual reductions in health care costs during the four-year Comprehensive Primary Care program while improving quality measures.²⁸

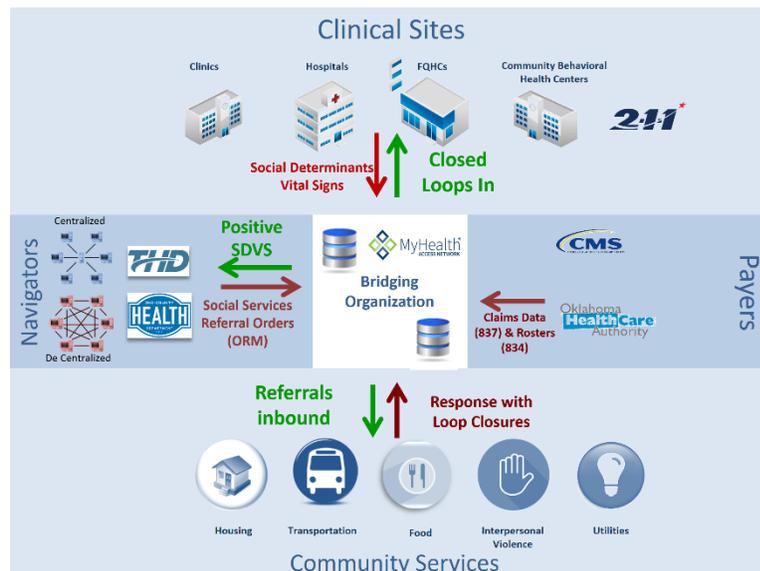
In the end, more than 90 percent of all savings in the seven-state demonstration project were achieved in Oklahoma. This work is now being expanded to address the social determinants of health and poverty such as food, housing, transportation, public utilities, and interpersonal violence through the Accountable Health Communities program (figure A.2).²⁹ MyHealth's Accountable Health Communities program, Route 66, screens individuals and families for social needs and provides navigators to connect them with needed services, tracking all data and monitoring for closed loops to assure high quality services delivery.

In addition, MyHealth is connecting data from early childhood education programs, and previously unknown early findings indicate that more than 45 percent of clients receive services in more than one organization.³⁰ Finally, MyHealth now serves as the Central Region hub for a nationwide health data-sharing initiative, the Patient Centered Data Home™, which is on track to connect health records for more than 200 million patients in America over the next few years.³¹ The architecture and policies of the Patient Centered Data Home network could provide a convenient scalability approach for regional efforts to quickly achieve nationwide reach.

By creating the Community Data Commons with attention to governance and stakeholder alignment, MyHealth has created new infrastructure that is now being leveraged by a wide variety of stakeholders in new and creative ways to improve health and well-being.

FIGURE A.2

Route 66 Program Infrastructure



Source: MyHealth.

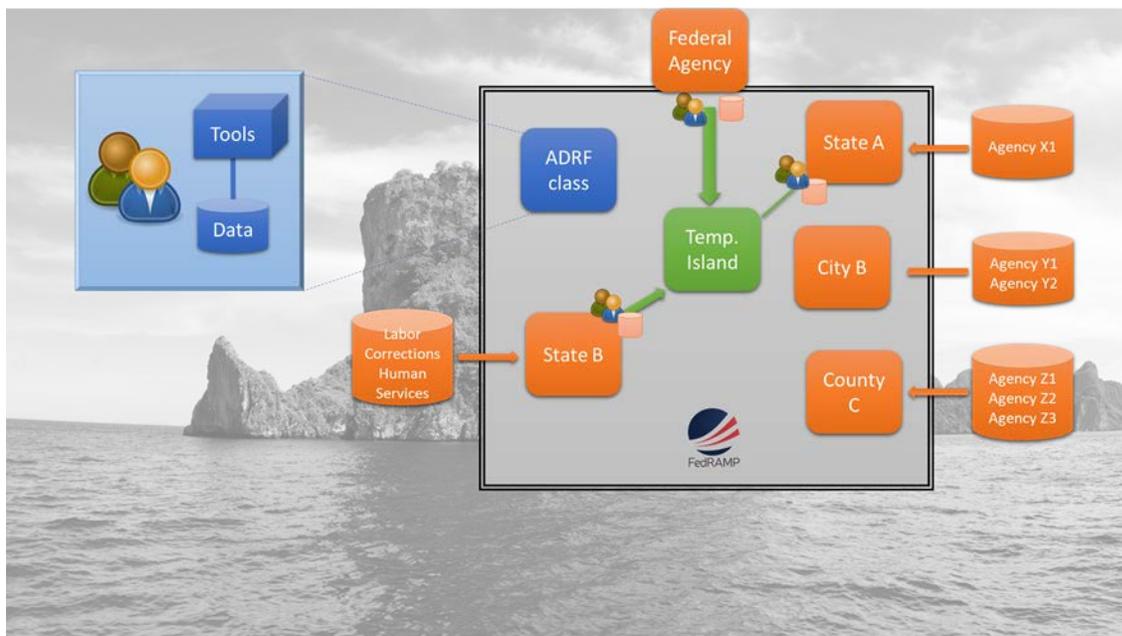
The Coleridge Initiative's Administrative Data Research Facility

The New York University Administrative Data Research Facility (ADRF) offers a new approach to sharing data across agencies and across state lines as part of its Coleridge Initiative (<https://coleridgeinitiative.org/>). It enables multiple local, state, and federal agencies to keep data in separate secure areas, approved by FedRAMP (<https://www.fedramp.gov/>) protocols, and link data for approved projects.

The ADRF allows agencies within the same state or different states to agree to share their data in a common area in the cloud for specific approved projects. If approved, staff from multiple agencies can access the common area, so they can work together to develop new integrated datasets, share information about coding differences or similarities, and develop common measures. No personally identifiable information is stored in the cloud; the data are hashed and deidentified (with a common hash algorithm) before being transferred to the secure area. Moreover, data stewardship modules can be deployed so agencies can track use and work output. More than 175 agency staff from about 50 agencies have already accessed and used the ADRF. Sensitive Title 13 Census Bureau data are also in the ADRF in a separate secure environment.

FIGURE A.3

A Cloud-Based Secure Environment That Allows Agencies to Share Data for Approved Projects



Source: The Coleridge Initiative, <https://coleridgeinitiative.org/computing>.

The ADRF was set up by the Census Bureau to inform the decisionmaking of the Commission on Evidence Based Policy (and was highlighted in the commission’s final report).³² It was designed to build on the lessons learned from the Census Bureau’s Local Employment Dynamics (and LEHD) program—mainly, that state and local access to and use of linked data were essential to the creation of high-value products. Access to the ADRF was made possible by the Laura and John Arnold Foundation and the Overdeck Family Foundation through scholarships to government agency staff.

After participating in a set of courses enabling staff from state agencies to work in the sandbox environment, agency staff were able to use linked data much more effectively. For example, they were able to use exit and admission data on formerly incarcerated people from the Illinois Department of Corrections, data on people receiving public benefits from the Illinois Department of Human Services, and unemployment insurance wage records from the Illinois Department of Employment Security to answer such questions as how neighborhood characteristics and transportation affect earnings, employment, and public assistance usage.³³

Notes

- ¹ See David T. Ellwood and Nisha G. Patel, *Restoring the American Dream: What Would It Take to Dramatically Increase Mobility from Poverty?* (Washington, DC: US Partnership on Mobility from Poverty, 2018).
- ² See Michael Holland, “Thoughts on the Organizational Structure of the Nationwide Data Initiative” (Center for Urban Science Progress, n.d.); Robert Grossman, “The Organizational Structure of the Nationwide Data Initiative: Considerations” (n.d.); and Andrew Reamer, “Nationwide Data Initiative: Principles of Approach to Organizational Design and Development” (George Washington Institute of Public Policy, 2018); all available at <https://coleridgeinitiative.org/resources>.
- ³ Robert M Goerge, “Barriers to Accessing State Data and Approaches to Addressing Them,” *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 122–37, doi:10.1177/0002716217741257; John Petrila, “Turning the Law into a Tool Rather than a Barrier to the Use of Administrative Data for Evidence-Based Policy,” *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 67–82, doi:10.1177/0002716217741088.
- ⁴ Michael Mueller-Smith, “Proposed Criminal Justice Administrative Records System National Data Schema” (Ann Arbor: University of Michigan, 2017).
- ⁵ Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin’s Press, 2018).
- ⁶ Jeffrey Mays, “Building an Infrastructure for Evidence-Based Policymaking: A View from a State Administrator,” *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 41–43, doi:10.1177/0002716217739275.
- ⁷ Commission on Evidence-Based Policymaking, *The Promise of Evidence-Based Policymaking* (Washington, DC: Commission on Evidence-Based Policymaking, 2017).
- ⁸ One example is SHIEC, the Strategic Health Information Exchange Collaborative, <http://strategichie.com>.
- ⁹ Erin Fries Taylor, Stacy Dale, Deborah Peikes, Randall Brown, Arkadipta Ghosh, Jesse Crosson, Grace Anglin, et al., *Evaluation of the Comprehensive Primary Care Initiative: First Annual Report* (Baltimore: Centers for Medicare & Medicaid Services, 2015); Deborah Peikes, Erin Fries Taylor, Stacy Dale, Ann O’Malley, Arkadipta Ghosh, Grace Anglin, Kaylyn Swankowski, et al., *Evaluation of the Comprehensive Primary Care Initiative: First Annual Report* (Baltimore: Centers for Medicare & Medicaid Services, April 2016); Deborah Peikes, Grace Anglin, Erin Fries Taylor, Stacy Dale, Ann O’Malley, Arkadipta Ghosh, Kaylyn Swankowski, et al., *Evaluation of the Comprehensive Primary Care Initiative: Third Annual Report* (Baltimore: Centers for Medicare & Medicaid Services, December 2016).
- ¹⁰ “Improving Outcomes for Young People at Risk—Transcript,” Stats NZ, last updated September 2017, http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/researchers-using-idi/better-data-better-lives/improving-outcomes-young-people-transcript.aspx.
- ¹¹ Alex Szalay, “From SkyServer to SciServer,” in “A Road-Map to Transform the Secure and Accessible Use of Data for High Impact Program Management, Policy Development, and Scholarship,” 2017, <http://policydatainfrastructure.com/index.html>; Holland, “Thoughts on the Organizational Structure of the Nationwide Data Initiative”; Grossman, “The Organizational Structure of the Nationwide Data Initiative: Considerations.”
- ¹² Andrew Reamer, Julia Lane, Ian Foster, and David Ellwood, eds., “Developing the Basis for the Secure and Accessible Use of Data for High Impact Program Management, Policy Development, and Scholarship,” special edition, *Annals of the American Academy of Political and Social Science* 675, no. 1 (January 2018).
- ¹³ Goerge, “Barriers to Accessing State Data and Approaches to Addressing Them.”
- ¹⁴ Daniel Goroff, Jules Polonetsky, and Omer Tene, “Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data,” *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 46–

- 66, doi:10.1177/0002716217742605; Simson L Garfinkel, "Privacy and Security Concerns when Social Scientists Work with Administrative and Operational Data," *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 83–101, doi:10.1177/0002716217737267.
- ¹⁵ Petrilá, "Turning the Law into a Tool Rather than a Barrier."
- ¹⁶ Dennis Culhane, John Fantuzzo, Matthew Hill, and TC Burnett, "Maximizing the Use of Integrated Data Systems: Understanding the Challenges and Advancing Solutions," *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 221–39, doi:10.1177/0002716217743441; Goerge, "Barriers to Accessing State Data"; Julia Lane, "Building an Infrastructure to Support the Use of Government Administrative Data for Program Performance and Social Science Research," *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 240–52.
- ¹⁷ The Health Insurance Portability and Accountability Act, or HIPAA, protects the privacy of medical records and personal health information. The Family Educational Rights and Privacy Act, or FERPA, protects the privacy children's education records; and 42 CFR Part 2 protects the confidentiality of medical information for people who receive or have received treatment for substance use disorder.
- ¹⁸ See <http://www.equality-of-opportunity.org/>.
- ¹⁹ "Moving to Opportunity (MTO) for Fair Housing Demonstration Program," <http://www.nber.org/mtopublic/>.
- ²⁰ Alexander S. Szalay, "From SkyServer to SciServer," *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 202–20; Amy O'Hara and Carla Medalia, "Data Sharing in the Federal Statistical System: Impediments and Possibilities," *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 138–50, doi:10.1177/0002716217740863.
- ²¹ Mays, "Building an Infrastructure for Evidence-Based Policymaking."
- ²² Jeffrey B. Liebman, "Using Data to More Rapidly Address Difficult U.S. Social Problems," *Annals of the American Academy of Political and Social Science* 675, no. 1 (2018): 166–81, doi:10.1177/0002716217745812.
- ²³ "Integrated Data Infrastructure," Stats NZ, last updated November 16, 2017, http://m.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx.
- ²⁴ Julia Lane and Tim Maloney, "Overview: The New Zealand Conference on Database Integration and Linked Employer-employee Data," 2002.
- ²⁵ "How Researchers Are Using the IDI," Stats NZ, last updated December 4, 2017, http://m.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/researchers-using-idi.aspx.
- ²⁶ "Multi-Agency Data Integration Project (MADIP)," Australian Bureau of Statistics, last updated March 22, 2018, <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Data+Integration+-+MADIP>.
- ²⁷ "Who We Are," MyHealth Access Network, accessed April 20, 2018, <https://myhealthaccess.net/who-we-are>.
- ²⁸ "Comprehensive Primary Care (CPC) Initiative: 2016 Shared Savings & Quality Results," Centers for Medicare & Medicaid Services, September 2017.
- ²⁹ "Accountable Health Communities Model," Centers for Medicare & Medicaid Services, last updated January 8, 2018, <https://innovation.cms.gov/initiatives/ahcm/>.
- ³⁰ David C. Kendrick, "Community Holistic Longitudinal Data System Stage 1 Results," presentation to George Kaiser Family Foundation Birth to Eight Strategy for Tulsa Program kick-off, September 7, 2017.
- ³¹ "What Is the Patient Centered Data Home?" Strategic Health Information Exchange Collaborative, accessed April 20, 2018, <http://strategichie.com/initiatives/pcdh/>.

³² Commission on Evidence-Based Policymaking, *The Promise of Evidence-Based Policymaking*.

³³ For sample projects, see the “Training” page on the Coleridge Initiative website, <https://coleridgeinitiative.org/training>.



www.mobilitypartnership.org