SEPTEMBER 2022

EDITH YANG

CDI | MDRC CENTER FOR DATA INSIGHTS

THE THREE INGREDIENTS OF SUCCESSFUL DATA PROJECTS

Data play a vital role in informing decision-making in organizations across all sectors and sizes. For example, in government agencies, administrative data are often collected and used for accounting, reporting, and compliance purposes, and program administrators are also increasingly viewing data as valuable information that can be used to improve program performance.[1] Many organizations, however, do not use their data to the data's full potential, which can directly affect how they approach delivering on their missions. Meanwhile, technology has developed in such a way that organizations, and society at large, can now access more powerful data tools, including complex artificial intelligence (AI) methods, to conduct and automate a variety of data-analysis methods for themselves, often in real time and on very large data sets. However, organizations can only make good use of these methods if they have access to data from related databases that are stored and maintained separately, and the fact that many of them do not have the means to integrate such databases or methods of good quality control continues to make it difficult for organizations to execute data projects successfully.[2] In fact, big data and AI data projects have a surprisingly high failure rate across public, private, and nonprofit organizations.[3]

The Center for Data Insights at MDRC has partnered with organizations to develop and execute a variety of data projects, some with long-standing partners that have participated in MDRC's long-term evaluations; others with newer partners that are just beginning to use data for program improvement. Important lessons about the essential ingredients for implementing successful data projects have emerged from this collaborative work. This brief discusses these ingredients in the context of the Center's [TANF Data Innovation (TDI)] project. TDI is a national initiative launched in 2017 by the Office of Planning, Research, and Evaluation and Office of Family Assistance (both within the Administration for Children and Families at the U.S. Department of Health and Human Services). The initiative helps Temporary Assistance for Needy Families (TANF) agencies enhance and build their ability to use their own administrative data to answer policy-relevant questions.[4] As part of this project, the Center partnered closely with eight state TANF agencies to design, launch, and complete data projects to address their analytical questions. The Center also worked with the Administration for Children and Families to design this initiative to address known challenges, and to develop projects that can inform how the federal program can support local TANF agencies. The lessons presented in this brief reflect a three-year collaborative experience, during which all project partners considered not only new data tools and skills, but also new ways of learning together.

The Center for Data Insights considers a data project a success when data are used routinely to inform policy and practice with the goal of improving the lives of the families served by programs. Box 1 lists examples of common questions workforce and education partners have asked and engaged the Center to help them answer using their data. The Center has learned from these partnerships that barriers to success are not primarily about technical issues or analytic methods. Rather, data projects need three ingredients to be successful: **people, perseverance,** and **project scoping** that includes a clear understanding of a project's needs and goals.

THE THREE INGREDIENTS: PEOPLE, PERSEVERANCE, AND PROJECT SCOPING

Guidance provided on how to launch data projects often focuses on technical details, programming languages and software, and data-analysis methods. But the three ingredients covered in this brief are more critical than data systems and technical expertise to set up the right conditions for organizations to learn from their data and improve.[5] This section uses the TDI partnerships as examples of how to make informed decisions in data projects by focusing on people, perseverance, and effective project scoping. These are not particularly novel concepts, but approaching them in a way that facilitates effective data use is more easily said than done.

People

The most central ingredient for a successful data project is people. Almost every data project involves people who provide data (often about themselves or their families), collect and analyze data (by importing data on people into data systems and by using statistical tools and methods to produce findings), and make decisions based on data (by reviewing the findings and taking action). Successful data projects require **human-centered design**, which focuses on the needs and expertise of these groups of people, beginning with those most affected by the projects. Historically, data-analytics

BOX 1

COMMON QUESTIONS IN WORKFORCE AND EDUCATION PROGRAMS THAT CAN BE ANSWERED WITH DATA

To begin addressing how to set up a data project for success, it is useful to gain a broad understanding of the common questions organizations want to answer and analytic methods that can be used to address them. The questions presented in the table are common ones that MDRC's workforce partners often ask. These questions determine the analytic methods or tools that project teams have used to answer them, with the goal of increasing programs' effectiveness or bringing programs to a larger scale.

QUESTION	ANALYTIC METHOD OR TOOL
What is the typical experience of a program participant during each stage of the program? Where are the significant points of attrition in the program's application, admission, or completion process?*	Journey mapping and funnel analysis
What are the most promising job clusters of employers or industries for program participants?†	Clustering
How should services be tailored to different segments of participants?‡ How do the impacts of services on employment outcomes vary among important subgroups?§	A/B testing and subgroup analysis
What are the different educational and employment trajectories taken by participants that lead to successful outcomes?‖	Success mapping
Who is most at risk of not reaching program milestones? Under what circumstances are they at risk?#	Predictive modeling
What are the different educational and employment trajectories taken by participants? How are participants' outcomes related to these trajectories?**	Trajectory modeling

*Edith Yang and Alissa Stover, "Putting Lived Experience at the Center of Data Science" (New York: MDRC, 2022).
†Maury B. Gittleman and David R. Howell. "Changes in the Structure and Quality of Jobs in the United States: Effects by Race and Gender, 1973–1990," *ILR Review* 48, 3 (1995): 420–440.
‡Washington Student Achievement Council, "Generous $1.5 Million Capital One Foundation Grant Will Fund Research to Support Postsecondary Education Opportunities for Low-Income Students in the State" (website: https://wsac.wa.gov/sites/default/files/2022.05.09.CapitalOneOtterbotGrant.pdf, 2022). This announcement provides a description of the Center for Data Insight's partnership with the Washington Student Achievement Council, which will explore how to direct services and resources to the students who most need them.
§Deondre' Jones, "Industry-Focused Training Has the Power to Reduce Inequities in Employment," *Workshift* (https://workshift.opencampusmedia.org/industry-focused-training-has-the-power-to-reduce-inequities-in-employment/, 2022).
‖Cynthia Miller, Victoria Deitch, and Aaron Hill, *The Employment Retention and Advancement Project: Paths to Advancement for Single Parents* (New York: MDRC, 2010).
#Camille Preel-Dumas, Richard Hendra, and Dakota Dennison, "Keep It Simple: Picking the Right Data Science Method to Improve Workforce Training Programs" (New York: MDRC, forthcoming).
**Kelsey Schaberg and Deondre' Jones, *Comparing Long-Term Employment and Earnings in Welfare Programs: Portland, Oregon, Early 1990s*, OPRE Report 2022-147 (Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, forthcoming).

projects undertaken by organizations that are meant for public good rather than for profit—including research organizations such as MDRC—have made the people who would be most affected by projects the subject of their work, but not always partners in that work. Being "human-centered" requires project partners to rethink and transform the way projects are designed and analyzed. It means including the perspectives of people the most affected by a project in all of its stages, from planning to completion, and it may require careful planning to invest in consultations with them.[6] It is easy get excited about applying new technologies and methods for analyzing data, so it is also important not to lose sight of the end goal of the analysis, which is to better meet participants' needs.

Another aspect of human-centered design has to do with project staffing**. Teams need to be interdisciplinary** to be relevant and successful. It is vital to incorporate people of diverse backgrounds and skill sets beyond data expertise. People with backgrounds in direct service provision, program administration, program operations, and other social or behavioral science disciplines bring contextual and applied knowledge that are important for those with technical expertise to understand, so that they can better streamline their work. The New York State Office of Temporary Disability Assistance found that team members who did not have technical expertise provided content expertise and contextual information that were instrumental for data-intensive processes such as data checking and cleaning, as well as interpreting the data findings. For example, nontechnical program staff members on the team who knew about program rules and policies for both unemployment insurance and TANF systems were able to help identify whether data inconsistencies between the TANF and wage-data systems were errors or expected outliers. The New York TDI team also sent nontechnical staff members to attend applied data analytics training, so that all members of the team could speak a common vocabulary when discussing data. Box 2 provides additional details about how the New York office built this interdisciplinary team, and how doing so led to better data work.

Finally, **data projects need strong support from leadership** to be successful. Having such support is particularly important for cross-disciplinary projects, since larger and more diverse teams mean more logistical challenges (for example, with scheduling or budgets). Leaders can also be instrumental in helping to communicate the importance of the project at the start, while the project is underway, and at the end when recommendations will need to be implemented.

Perseverance

Perseverance is essential to success in any endeavor. Data projects in particular go through twists and turns as teams face issues of data access, data limitations, and data quality. Often the data issues that must be addressed before any analysis can occur take much longer to examine and resolve than planned. Since unanticipated challenges are inevitable in data projects, those projects must expect failures on the path to success and use them as learning opportunities. A project's milestones, or even its purpose, may also change with these internal or external shifts.

In data analytics, teams using new approaches and techniques often go through a cycle in which they have high expectations for exciting results from a planned data analysis, experience an analytic

BOX 2

THE NEW YORK STORY: USING OPERATIONAL, SUBSTANTIVE, AND LIVED EXPERTISE TO INCREASE THE USE OF DATA FOR PROGRAM IMPROVEMENT

The New York State Office of Temporary and Disability Assistance was interested in helping local TANF districts identify, early on, participants who might become long-term recipients of public assistance benefits. Before it joined the TDI project, this office (which includes the state's TANF program and a "safety net" program for families who are not eligible for the state's TANF program) collaborated across functions in an ad hoc way. Teams with distinct types of expertise usually worked independently. To pursue this predictive analytics project, the office brought together a cross-functional team that included 11 staff members from four different departments. That team included researchers, programmers, employment experts, and operational staff members who worked regularly with local districts. They met monthly and attended TDI's teaching and learning webinars. The team established clear roles and clear goals, encouraged communication and cooperation, and got explicit support from executive leaders, which led to increased team engagement.

Working in this way presented some challenges. People had to make time in already busy workdays for this in-depth collaboration. They had to communicate more often and openly. Different perspectives sometimes led to slower decision-making. At the same time, the team recognized that incorporating different perspectives also led to better decisions. People from different teams educated each other and learned how to speak a common language when communicating about data sources, data quality, and findings. All team members became conversant in the process of generating useful evidence to answer a question using the data they had available.

The team also developed ways to set future projects with diverse teams up for success. They launched a communication plan and created a culture of open and constant communication. They focused on creating trust among the different groups and embraced educating and learning from each other.

Building this culture of communication and trust allowed the team to ask deeper questions about the data and the analysis. Team members who did not usually work with or look at data learned from the data team how to run some basic analyses in Excel. The technical staff started to ask questions such as "Will this work?" and "How can different local offices use this information?" in addition to "Is this right or wrong?" when putting together their predictive models. The New York story highlights how including a wider range of people and drawing on those people's strengths can produce better-informed research findings that advance an organization's mission.

challenge that does not give them valid results, and then need to adjust their original plan. They might go through this cycle a few times before they start experiencing productivity.[7] When the team encounters a problem or limitation, it is important to lean into the problem and fully understand how it affects the rest of the project. Some issues can be resolved by changing analytical techniques or collecting cleaner data, while other issues may require a transparent accounting of how the issue limits the project's potential findings, or how the methods used may produce biased findings for certain groups of participants. The Colorado Department of Human Services experienced several rounds of data limitations and analytical challenges, but because it persevered through these challenges, its data project yielded new insights about the usefulness of its supportive payments program. Box 3 highlights that journey of perseverance.

BOX 3

THE COLORADO STORY: OVERCOMING DATA LIMITATIONS TO IMPROVE LIVES

The Colorado Department of Human Services implemented a data project that is a story of persistence and perseverance. When it joined the TDI project, it was interested in exploring how the timely provision of supportive payments for housing, transportation, and childcare affected employment for families on its TANF caseloads. The Colorado TDI team could use data from the agency's benefits-management system and unemployment insurance wage data from the Colorado Department of Labor and Employment to conduct their analysis.

Because the benefits-management system had only limited documentation on reasons people were receiving supportive payments, the team had to adjust their original research question and approach. Instead of focusing the analysis on one specific type of supportive payment, the team decided to look at all supportive payments, regardless of type. In addition, instead of looking at the timeliness of payments, they decided to simply look at the effect of receiving any payments. The team remained dedicated to discovering whether supportive payments were helpful to the most vulnerable families.

The plan was to compare the employment rates of parents who received supportive payments with those of parents who did not, using a regression model to adjust for differences between the two groups. But when the team ran the model, they encountered counterintuitive results. Looking at demographic characteristics of the two groups, they discovered **selection bias** issues; parents who received payments came from very different situations than parents who did not receive payments, in terms of characteristics such as household sizes or education levels. They worried that these differences were leading to misleading results about the usefulness of supportive payments.

Working closely with coaches from the Center for Data Insights to plan and conduct their analysis, the Colorado team then shifted to using **propensity score matching** to address selection bias. This method takes the group of parents who received supportive payments and "matches" each parent with one who did not receive a payment but had similar characteristics and experiences (for example, someone of the same gender, age, and level of educational attainment). Comparing these two more similar groups could give the team more confidence that differences in employment outcomes could be attributed to the supportive payments.

Learning a new analysis method was no small task—it took a significant investment of time for the Colorado team to specify a human-centered research question, assess the quality of the data, adjust their question to one they could answer, consider analysis methods that proved invalid, and continue to learn how to conduct an analysis they had limited experience with. Propensity score matching itself is an iterative process—the team designed and tried four different matching methods that ran into roadblocks before persevering to find one that worked.

Finally—after more than a year of failing often, adjusting, and overcoming numerous challenges—the Colorado team learned from their data that parents who received supportive payments were more likely to enter employment and maintain stable employment than parents who did not receive payments. They also learned that there was much more to learn and have plans to continue both supplementing their analyses and improving their agency's data-collection and -documentation practices so that they can better support Colorado families who rely on TANF services.

Project Scoping

Finally, effective project scoping lays the foundation for a successful data project. Project scoping involves laying out a detailed plan of action and timeline for launching, executing, and completing a project. Doing so effectively requires a clear understanding of the purpose of a project and the processes, tools, and expertise it will need to reach its learning goals. It means planning it as a true learning partnership that elevates the expertise of nontechnical or nonresearch partners. It means anticipating that initial assumptions about data, program operations, or program participants might prove to be inaccurate, and that the project might have to make changes in design or process. It also means making sure that consultations with the right people—including those with lived experience of the programs being analyzed—start early and will continue throughout the life of the data project. Scoping a project effectively is necessary for the other two ingredients to be present: It makes it possible for data projects to include all important perspectives and gives project teams time and resources to persevere through challenges.

The first step in effective project scoping is to **establish the purpose of your data project, making sure its goals are clear and the research questions are well defined**. State TANF agencies operate under different rules, have different levels of access to certain types of administrative data, and have different enforcement mechanisms. But they all share a goal of better serving TANF program participants in their states. The TANF agencies participating in the TDI project developed research questions that identified the people they would study and the intervention or circumstance that is potentially affecting them. For example, the New York State TDI team wanted to identify factors associated with long-term TANF receipt, and the Colorado TDI team was interested in the efficacy of providing supportive payments to program participants as a means of improving employment outcomes.

After establishing a clear purpose and research questions, teams can **develop a big vision but start executing with a small first step**. Organizations might start their data projects by prematurely upgrading their data systems or purchasing new ones, providing intensive data training, or hiring large data teams. But the TDI project's needs assessment found, counterintuitively, that TANF organizations viewed as "data leaders" tended to have *older* systems, suggesting that it was not critical for them to modernize their systems.[8] And, as discussed in the New York State TDI example, having interdisciplinary teams that include people with nontechnical expertise can improve data quality and provide the proper context for interpreting findings. Making sure that investments are aligned with the project's purpose and vision will help ensure that decisions made along the way are achievable and represent progress toward the project's goals.

Designers of data projects should **make room to be agile**.[9] Issues and challenges often arise while a project is in process. Effective project scoping can build in options for flexibility and anticipate the need for large and small changes. Teams often try to include every possible analysis and outcome in a data project from the outset. Planning to address all project questions or outcomes of interest comprehensively at each step of the process might be the most straightforward way to think about entering into data work, but such exhaustive planning can stall the progress of data collection, data cleaning, and exploratory analysis. It may be more effective to take a more agile project-management approach of focusing first on a simple analysis with a few data fields and working through the entire data process before adding more complexity to the project. Completing a simple analysis can help a team identify the challenging points in a data workflow, adjust processes to make that workflow smoother, and anticipate the levels of effort needed when more complexity is added to the project. That complexity could include creating measures from less familiar data sources or employing a more complicated statistical tool. Each iteration of the analysis can then build on the lessons learned from the previous iteration and can yield further improvements.

Finally, teams need to scope data projects **realistically**. Data projects are often too ambitious—sometimes because some crucial perspectives are missing from the planning process. For example, if experienced analysts are not involved in project scoping, teams might skip over basic descriptive analysis—which looks at simple proportions and averages and can yield useful insights about the people programs serve—and jump to causal questions that require rigorous procedures to be answered properly. The Colorado TDI pilot team's data-analysis journey, as described in Box 3, was full of twists and turns resulting from important lessons learned from their initial descriptive analyses. Additionally, if teams overlook guidance from people with lived expertise, they might plan complex analyses that they find exciting but may yield less useful findings for the field with methods that may be harder to perform or replicate.

CONCLUSION

Data projects are often exciting to envision but challenging to execute. Despite the boom in data science, projects that involve large data sets and powerful data tools still have a surprisingly high failure rate.[10] But projects can succeed if they take people-centered perspectives, persevere through analytic challenges, and are scoped effectively. As the highlighted stories show, setting up data projects for success is not primarily about data or data systems. It is about people—planning, designing, and pushing through challenges together. Projects that are scoped effectively, focus on the people affected, and encourage teams to persevere through challenges yield better results and richer data findings and lessons, and ultimately lead to better ways for organizations such as state and local TANF agencies to fulfill their mission of improving people's lives.

NOTES AND SOURCES

1 Melissa Wavelet, Stephanie Rubino, Caroline Morris, and Maryah Garner, *Increasing Data Analytics Capacity in State TANF Agencies: The TANF Data Collaborative Approach*, OPRE Report 2022-124 (Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022).

2 IBM Institute for Business Value, "Dealing with the AI Data Dilemma" (website: https://www.ibm.com/downloads/cas/EBJQ6K7M, 2021).

3 Nick Hotz, "Why Big Data Science & Data Analytics Projects Fail" (website: https://www.datascience-pm.com/project-failures, 2022).

4 Administrative data are those the agencies are already collecting primarily for the management of their programs and services.

5 Robert M. Goerge, Emily R. Wiegand, Emma K. Monahan, and Leah Gjertson, *Exemplary Data Use by State TANF Agencies*, OPRE Report No. 2022-133 (Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022).

6 Edith Yang and Alissa Stover, "Putting Lived Experience at the Center of Data Science" (New York: MDRC, 2020).

7 Jackie Fenn, "Understanding Gartner's Hype Cycles, 2007," Gartner ID G00144727 (Stamford, CT: Gartner, Inc.).

8 Goerge, Wiegand, Monahan, and Gjertson (2022).

9 Ken Schwaber, *Agile Project Management with Scrum* (Redmond, WA: Microsoft Press, 2004).

10 Hotz (2022).

ACKNOWLEDGMENTS

The author and the TANF Data Innovation (TDI) coaches at the Center for Data Insights would like to thank the Center's TDI partners, especially the New York and Colorado pilot teams, who allowed us to share their data stories and whose expertise, dedication, and care for the residents in their states propelled their data projects. From the New York State Office of Temporary and Disability Assistance, the interdisciplinary and cross-functional team included Stephanie Boshart, Jerrod Ogden, Bill Koncelik, and Liz Morrisey from Employment and Advancement Services; Alison Maura, Shannon Al-Jabi, and Melissa Alexander from the Temporary Assistance Bureau; Renee Nowicki from the Division of Budget, Finance, and Data Management; and Dave Dlugolecki, Britany Orlebeke, Fang Cheng, Solanlly Cabrera, and Lauren Springer from the Bureau of Data Management and Analysis. Megan Kauffman, Michael Martinez-Schiferl, and John McGraw from the Colorado Department of Human Services designed and executed the Colorado pilot project. We appreciate Girley Wright, Nicole Deterding, and Lauren Frohlich, our sponsors and partners at the Administration for Children and Family's Office of Family Assistance and Office of Planning, Research, and Evaluation, for their generous support of the TDI project and for making this brief possible. Additionally, many thanks to our colleagues Richard Dorsett, Erika Lundquist, and Johanna Walter, who collaborated with the TDI teams on their journeys, and Eva Ben-Schlomo, Richard Hendra, John Hutchins, David Selby, Melissa Wavelet, and Donna Wharton-Fields, who were instrumental in inspiring, shaping, and reviewing early drafts of this brief. Finally, Joshua Malbin edited the report, and Ann Kottner prepared it for publication.

Dissemination of MDRC publications is supported by the following organizations and individuals that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, Arnold Ventures, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, and Sandler Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

NEW YORK
200 Vesey Street, 23rd Flr., New York, NY 10281
Tel: 212 532 3200

OAKLAND
475 14th Street, Suite 750, Oakland, CA 94612
Tel: 510 663 6372

WASHINGTON, DC
750 17th Street, NW, Suite 501
Washington, DC 20006

LOS ANGELES
11965 Venice Boulevard, Suite 402
Los Angeles, CA 90066