A TOOLKIT FOR

Centering Racial Equity Throughout Data Integration

VERSION 2.0

AISP

Societal “progress” is often marked by the construction of new infrastructure that fuels change and shifts the landscape. Just as railroads and interstate highways were the defining infrastructure projects of the 1800s and 1900s, the development of data infrastructure is a critical innovation of our century.

Railroads and highways were drivers of development and prosperity for some. Yet other individuals and communities were harmed, displaced, bypassed, or ignored by those efforts.

As railroads and highways both developed and decimated communities, so too can data infrastructure.

At this moment in our history, we can co-create data infrastructure to promote racial equity and the public good, or we can invest in data infrastructure that disregards the historical, social, and political context—reinforcing racial inequity that continues to harm communities. Building data infrastructure without a racial equity lens and understanding of historical context will exacerbate existing inequalities along the lines of race, gender, class, and ability. Instead, we commit to contextualize our work in the historical and structural oppression that shapes it, and organize partners across geography, sector, and experience to center racial equity in data infrastructure.

Author's Note

We, the authors of the toolkit, have been working on this updated version for nearly two years. In that time, we have poured over each sentence and considered every word carefully. This work is a labor of love for everyone involved. It was created in the spirit of bettering ourselves, bettering the work we do, and bettering the lives of all people, but especially those harmed by systemic injustice. In 2020, when the original toolkit was released, it seemed that collective consciousness across the United States was shifting in this direction, that we were meeting the culture at a critical inflection point and pushing forward together. Now, in February 2025, releasing this work feels defensive. We are pushing against seemingly enormous odds, trying to mitigate a barrage of harms.

At AISP, we sit in a position of relative privilege, housed at a university and funded primarily philanthropically. Given this position, we have chosen not to change the language in this toolkit. We firmly believe that diversity, equity, and inclusion are important. We believe that disparate impacts based on race, ethnicity, sexual orientation, gender, and disability exist and can be either replicated or mitigated by data use. We believe in the humanity of every person represented by a data point.

We have removed the names of several contributors and anonymized some Work in Action examples to protect those who might be put at risk by being associated with this work. We are so grateful for their contributions and are strengthened by the knowledge that they continue to embody equity despite a prohibition on the word.

We have decided that these chaotic first six weeks of the new administration will not override two years of hard work. Speaking plainly and speaking truthfully are radical acts. We choose to use these words for all of those who cannot. Thank you for accompanying us on this journey.

In solidarity and hope,
Actionable Intelligence for Social Policy

As you move through the toolkit, remember:

"Those of us committed to racial justice, democracy, and mutual well-being know that safety does not come from hatred; it comes from solidarity with all who yearn for a world centered on mutual care and concern."
— Tema Okun

Introduction

Today, vast amounts of data are collected by public agencies, and data infrastructure and use are expanding at every level of government—local, state, and federal. There are positive aspects to this: Data have the potential to help communities illuminate disparities, hold governments accountable, and support collaborative action. Yet, considerations around racial equity are often absent from data infrastructure efforts, and tools to utilize and mine these data are expanding exponentially. This is especially troubling because government policies and programs that produce administrative data have often played a direct role in creating, enabling, and sustaining institutional and structural racism.

We aim to change this.

With trust in government and "experts" at historic lows,[1] efforts that rely exclusively on these institutions to "use data to solve social problems" are unlikely to succeed. Too often, government organizations and their research partners fail to identify and address issues of bias in data. Further, even if such issues are identified, these organizations are not equipped to repair trust with communities that have experienced harm.

We envision data infrastructure and use as a means to confront racism, expose injustice, act on our shared values, and elevate lived experience.

This body of work seeks to encourage shifts of awareness and practice by centering racial equity and community voice as we build and remake cross-sector data infrastructure. Our vision is one of ethical data use that allows us to understand and address human needs holistically, and supports information and power sharing with community members.

Key Terms

- **Racial equity** is a process of eliminating racial disparities and improving outcomes for everyone. It is the intentional and continual practice of changing policies, practices, systems, and structures by prioritizing measurable change in the lives of people of color. (See Race Forward.)
- **Data infrastructure** refers to the systems, technologies, and processes for using, storing, securing, and interpreting data. This includes hardware, software, and organizational practices.
- **Administrative data** are data collected during the routine process of administering programs.
- **Administrative data reuse** involves using these data in a way not originally intended (e.g., for research).
- **Data sharing** is the practice of providing access to information not otherwise available.

---

[1] Pew Research Center. (2024). Public Trust in Government 1958-2024.

- **Data integration** is the process of bringing together data from different sources, which often includes identifiable information (e.g., name, date of birth, SSN) so that records can be linked at the individual level.

Data sharing and integration involve significant privacy risks, and all data use should be carefully considered to ensure that sharing is legal and ethical, with a purpose that can be linked to action.

Why Data Integration + Racial Equity?

Cross-sector data sharing and integration enable the use of multiple sources of information to better understand individual, family, and neighborhood experiences and conditions over time. With more cross-sector data, we can often better capture both the causes and impacts of complex social issues and improve programs, policies, and funding approaches to build stronger communities. **Yet, the way that cross-sector data are used can also reinforce legacies of racist policies and produce inequitable resource allocation, access, and outcomes.**

We understand **structural racism** as the normalization and legitimization of historical, cultural, institutional, and interpersonal dynamics that advantage Whites, while producing cumulative and chronic adverse outcomes for people of color. Embedded within structural racism is **institutional racism,** the ways policies and practices of organizations or parts of systems (schools, courts, transportation, etc.) create different outcomes for different racial groups (see Terms).

Black, Indigenous, and people of color as well as people living in poverty are often over-represented within government agency data systems, and disparate representation in data can cause disparate impact.[2] Laws, policies, business rules, and narratives are permeated by structural racism, which is the root cause of the racial disparities evident in system outcomes. Such disparities are often sterilized by well-intentioned names (e.g., "disproportionate contact" in the legal system or the "achievement gap" in education) that hide the social consequence of structural racism: that, as a group, Black, Indigenous, and people of color in the United States have worse outcomes in many human service system outcome measures regardless of socioeconomic status.[3] And yet, many agency solutions and data initiatives are largely disconnected from this root cause, and the "hunt for more data is [often] a barrier for acting on what we already know."[4]

---

[2] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, *104*, 671.

[3] Hayes-Greene, D., & Love, B. P. (2018). *The Groundwater approach: Building a practical understanding of structural racism.* Racial Equity Institute.

[4] Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Wiley.

With this knowledge, we call for users of administrative data and those building data integration capacity to center racial equity in their data practices. We call for the inclusion of community voices and power sharing at every stage of the data life cycle. We call for relationship building among those represented in the data and those stewarding and using the data. Without a deliberate effort to address structural racism, institutional racism, and unrecognized bias, data integration will inevitably reproduce and exacerbate existing harm.

To avoid this, we must embed considerations of racial equity throughout the data life cycle:

- In planning
- In data collection
- In data access
- In data analysis
- In the use of algorithms & artificial intelligence
- In reporting & dissemination

We are at a pivotal moment, one in which the use of data is accelerating in both exciting and concerning ways. While we have access to greater amounts of data than at any other point in our history, privacy laws and data governance practices lag behind, placing Black, Indigenous, and communities of color at the greatest risk of the "data-ification of injustice."[5]

Acknowledging history, harm, and the potentially negative implications of data integration for groups marginalized by inequitable systems is a key first step, but it is only a first step. To go beyond this, we must center the voices, stories, expertise, and knowledge of these communities in decision-making, and take collective action with shared power to improve outcomes and harness data for social good.

We are working to create a new kind of data infrastructure—one that dismantles "feedback loops of injustice"[6] and instead shares power and knowledge with those who need systems change the most.

Will you join us?

---

[5] Benjamin, R. (2019)

[6] Eubanks, V. (2018). *Automating Inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Contents

Who We Are

Actionable Intelligence for Social Policy (AISP) is based at the University of Pennsylvania within the School of Social Policy & Practice. We support the ethical use of individual-level administrative data for social policy change and advocate for the expansion of resources and infrastructure that makes this possible. We foster cross-sector collaborations, build the relationships and trust that enable and sustain data sharing, and center racial equity in our approach to data sharing and integration. We are the conveners of the Workgroups that guided development of this Toolkit.

Toolkit contributors are a diverse group of civic data enthusiasts, including community advocates, staff of local and national nonprofit organizations, applied researchers, state and local government administrators and analysts, foundation staff, and service providers. The first version of this Toolkit was published in May 2020 after a two-year participatory process. As with most long-term projects, we saw a need for revisions immediately following release. This updated version incorporates threads of feedback from the hundreds of practitioners we have had the opportunity to learn alongside while presenting and discussing this work. We began convening a new Workgroup in 2024, with many contributors serving consistently from 2018 to 2025. Thank you to all who have engaged with us on this journey!

This Toolkit sits at the intersection of dynamic yet often disconnected fields—building civic data infrastructure to inform decision-making and the work for racial equity. Both fields have had dramatic shifts in the past five years, and the content of this Toolkit has also changed. We have listened to you as you told us how you use this resource. We have modified parts that were less helpful, shortened some sections, and expanded others. Most notably, we have incorporated essential topics that were missing—guidance and Work in Action specific to participatory governance, Tribal data sovereignty, and data standards. We have also updated terms and language used to describe these ever-changing topics. You may also notice that artificial intelligence has been incorporated as a part of the data life cycle. All changes have come from users of this Toolkit, and we are grateful for your help in moving the field forward.

Together, we have worked to co-create strategies and identify best practices for administrative data integration and reuse by government agencies, nonprofit organizations, and data collaboratives committed to centering racial equity and sharing power with community.[7] Specifically, our work seeks to help agencies acknowledge and compensate for the harms and bias baked into public data, data structures, and data practices, as well as the perceptions and narratives they fuel.

2024–2025 AISP CONTRIBUTORS

Amy Hawn Nelson
Sharon Zanti
Della Jenkins
Isabel Algrant
Jessie Rios Benitez
Emily Berkowitz
TC Burnett
Kristen Egoville
Deja Kemp
Blu Lewis
Kim Paull

---

[7] Throughout this toolkit we use the term 'we' to promote inclusivity. All contributors and readers bring unique experiences to this work, and while we do not purport to encompass all perspectives (because that is impossible), our intention is inclusion. The "we" includes all of us working at this intersection of data use and equity.

2024–2025 WORKGROUP CONTRIBUTORS

Tess Abrahamson-Richards, Hummingbird Indigenous Family Services
Meeta Anand, The Leadership Conference on Civil and Human Rights
Bridget Blount, Baltimore's Promise*
Angela Bluhm, Oregon Department of Education
Kathryn Darnall Helms, Oregon Enterprise Information Services
Jenrose Fitzgerald, Washington University in St. Louis
Robert Gradeck, Western Pennsylvania Regional Data Center*
Jordan Harvill, Advance Native Political Leadership
Sherri Killins Stewart, Build Initiative
Blu Lewis, NC BLOC & Our Data Bodies*
Brandeis Marshall, DataedX
Marjorie McGee, Oregon Health Authority
Saundra Mitrovich, Native American Rights Fund
Kim Paull, Blue Cross Blue Shield of Rhode Island*
Tawana Petty, Petty Propolis
Meenakshi Rao, Oregon Enterprise Information Services
Enid Rey, Brave Consulting Solutions*
Sumit Sajnani, Connecticut Office of Health Strategy
Sonia Torres Rodriguez, Urban Institute
Rebecca Whitaker, Duke University
Emily Wiegand, Chapin Hall
Thelma Wong, Jeong Collective*
Hayley Young, North Carolina Department of Health and Human Services
**Equity in Practice Fellow*

2024–2025 WORK IN ACTION CONTRIBUTORS

Albuquerque Area Southwest Tribal Epidemiology Center: Kevin English
Allegheny County (PA) Department of Human Services: Geoffrey Arnold, Kathryn Collins, & Ellen Kitzerow
Allegheny County (PA) Office of the County Manager: Joanne Foerster
Allied Media Project: Chiara Galimberti, Molly Leebove, & Toni Moceri
Asemio: Aaron Bean & Erin Powell
Atlanta Wealth Building Initiative: Jarryd Bethea & Alex Camardelle
Baltimore's Promise: Bridget Blount & James Sadler
Black Data Wealth Center: David Asiamah
Black Researchers Collective: Gleanace Green & Shari Runner
Boston University Center for Antiracist Research: Dawna Johnson & Ibram Kendi
Boston University School of Social Work: Dolores Acevedo-Garcia & Clemens Noelke
Camden Coalition: Taylor Brown & Martiza Gomez
Center for Public Sector AI: Justin Brown, Cassandra Madison, & Kristen Tillett
City of Boston, Information and Technology: Santiago Garces & Alejandro Jimenez Jaramillo
Community Solutions: Adam Reuge
Connecticut, Office of Early Childhood: Chanae Russell & Elena Trueworthy
Connecticut, Office of Policy and Management: Katie Breslin, Scott Gaul, & Pauline Zaldonis
Data You Can Use: Victor Amaya
Delaware, Department of Education: Caitlin Gleason
Embrace Boston: Jenny LaFleur
Hartford Data Collaborative (CT): Kate Eikel & Michelle Riordan-Nold

Indiana Management and Performance Hub: Josh Martin
Iowa State University, I2D2: Heather Rouse
Kentucky Center for Statistics: Matt Berry
Mecklenburg County (NC) Community Support Services: Mary Ann Priester
Metriarch: Jacqueline Blocker
MimiOnuoha.com: Mimi ỌNỤỌHA
Neighborhood Nexus:Nikolai Elneser & Tommy Pearce
Northside Achievement Zone: Amy Susman-Stillman
Open Data Charter: Cat Cortes
Philadelphia Monument Lab: Paul Farber
Seattle Information Technology Department: Jim Loter
State of Connecticut, Office of Health Strategy: Sumit Sajnani
State of Oregon, Oregon Health Authority, Oregon Behavioral Health Survey: Renee Boyd & Kusuma Madamala
The Folded Map Project: Tonika Lewis Johnson
UNC Chapel Hill, School of Information and Library Science & NC Data Works: Alex Chassanoff
University of Arizona:Jason Jurjevich
University of California-Los Angeles, California Policy Lab: Janey Rountree
University of Massachusetts-Amherst, School of Public Health and Health Sciences: Daniel Lopez-Cevallos
University of Michigan, Population Studies Center: Jordan Papp & Chandler Rombes
Western Pennsylvania Regional Data Center: Robert Gradeck & David Walker
University of Southern California, Children's Data Network: Regan Foust
Urban Institute: Claire Bowen, Leah Hendey, Gabe Morrison, Madeline Pickens, & Alena Stern
Wilder Research: Jessie Austin O'Neill, Briellen Griffin, Nicole MartinRogers, & Piere Washington

> **As of January 2025, several longtime contributors to this work decided to remain unnamed. We appreciate their contributions and regret the circumstances that threatened their ability to contribute publicly.**

Acknowledgments

The authors wish to express their gratitude to the Robert Wood Johnson Foundation and the Ford Foundation, as well as the other funders who supported this work. The conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funders.

We are grateful to the 2023-2024 students of the University of Pennsylvania's School of Social Policy & Practice course, Applied Data Ethics, Law, & the Social Good, who provided valuable critique and suggested updates, including Ryann Carey for her work sourcing Work in Action. We are also grateful to Gilly Gehri for their leadership in framing the Algorithms & Artificial Intelligence section as a part of an independent study with AISP on this topic. We also recognize Briellen Griffin, who provided an external review of this Toolkit.

Finally, we'd like to thank our 2020 AISP Contributors, Workgroup Contributors, and Work in Action Contributors who supported this work in 2019 and 2020 (names below).

2019–2020 WORKGROUP CONTRIBUTORS

Niiobli Armah, My Brother's Keeper, Equity Intelligence Platform
Bridget Blount, Baltimore's Promise
Angela Bluhm, Chief Education Office, State of Oregon
Katy Collins, Allegheny County Department of Human Services
Sheila Dugan, GovEx, Johns Hopkins University
Laura Jones, Writer and Community Advocate based in Minnesota
Chris Kingsley, Annie E. Casey Foundation
Ritika Sharma Kurup, StriveTogether
Tamika (Blu) Lewis, Our Data Bodies
Rick Little, Utah Dept of Human Services, Management Information Center
Tawana Petty, Detroit Community Technology Project & Our Data Bodies
Raintry Salk, Race Forward and Government Alliance for Racial Equity (GARE)
Michelle Shevin, Ford Foundation

2020 WORK IN ACTION CONTRIBUTORS

Allegheny County (PA), Department of Human Services, Office of Analytics, Technology, & Planning: Samantha Loaney, Brian Bell, Ellen Kitzerow, Julia Reuben, Shannon Flynn, & Jamaal Davis
Allegheny County (PA) Department of Human Services, Office of Equity & Inclusion:
Shauna Lucadamo & Jessica Ruffin
Automating.NYC: Deepra Yusuf, Elyse Voegeli, Akina Younge, & Jon Truong
Birth through Eight Strategy for Tulsa (BEST), Jessica England & Dan Sterba
City of Asheville (NC), Christen McNamara & Kimberlee Archie
City of Tacoma (WA), Alison Beason
DataWorks NC, Libby McClure & John Killeen
Kentucky Center for Statistics, Jessica Cunningham
Mecklenburg County (NC) Community Support Services, Courtney LaCaria & Mary Ann Priester
New York City Administration for Children's Services & Youth Studies Programs at the CUNY School of Professional Studies, Sarah Zeller-Berkman
Take Control Initiative (OK), Emma Swepston, Laura Bellis, & Brandy Hammons

2020 AISP CONTRIBUTORS

Amy Hawn Nelson
Della Jenkins
Sharon Zanti
Matthew Katz
Emily Berkowitz
TC Burnett
Dennis Culhane

Suggested Citation:

Hawn Nelson, A., Zanti, S., Jenkins, D., Algrant, I., Rios Benitez, J., et al. (2025). A Toolkit for Centering Racial Equity Throughout Data Integration. Actionable Intelligence for Social Policy, University of Pennsylvania. https://www.aisp.upenn.edu/centering-equity/

How to Use This Toolkit

This Toolkit is designed to help guide partnerships, collaboratives, agencies, and community initiatives seeking to center equity while using, sharing, and integrating administrative data.

> Not sure what we mean by using, sharing, and integrating administrative data? Take some time to review our Introduction to Data Sharing & Integration, which covers key terms, concepts, and first steps.

Are you interested in deepening your organization's work in centering racial equity throughout the data life cycle? We have created Your Journey to Centering Racial Equity: A Companion Workbook to this Toolkit. This Companion Workbook includes activities to help guide individuals, teams, and organizations in this work.

Some caveats: We use the data life cycle as an organizing principle, but we want to be clear that no matter where you are in the work, the goal is most often to take note of what you are doing and why, adjust speed (pause or get some urgency), and make adjustments in partnership with community. While we organize Work in Action examples using the data life cycle, these are somewhat arbitrary classifications, as examples often involve work across multiple stages.

Most importantly, this Toolkit is not meant to be prescriptive, comprehensive, or even read cover to cover. Users have told us that they most effectively use the Toolkit as inspiration, as a list of possible strategies rather than implementation guidelines. All of us have work to do in centering racial equity, but we will never move forward by trying to do everything at once. Please take what is helpful, and leave what is not.

Who Should Use This Toolkit

We believe that all voices are needed in conversations about racial equity and data use, and the information presented here can be used by anyone. This Toolkit and Companion Workbook are specifically framed to support civic data users in their efforts to center racial equity.

These users could include:

- **Members of institutions:** university-based researchers; government-agency administrators and analysts; foundation staff
- **Advocates and community members:** community and religious leaders; civic and neighborhood association members; students, parents, and caregivers
- **Bridges between community and organizations:** service providers; social workers; teachers; local backbone organizations; independent applied researchers

Many types of civic data use are relevant to this Toolkit, including:

- Open data (data that can be shared without legal agreements in place)
- Protected administrative data (confidential data that can be shared only with sufficient security provisions in place, including data sharing agreements)
- Integrated data systems (systems that regularly link protected administrative data across agencies to improve evidence-based collaboration)
- Dashboards (administrative data aggregated to topic/indicator/subgroup/population)
- Neighborhood indicators (data aggregated to place)
- Research, evaluation, and outcome measurement using administrative data
- Tools created by using administrative data, such as algorithms

Foundations for Community Involvement

Assessing Readiness

This Toolkit aims to support you in creating a new kind of data infrastructure that shares power with community. The reality is that getting there is a long and winding road that we each travel in distinct contexts. If we jump into conversations with community members before we have examined our individual and institutional readiness, we risk doing more harm than good. Readiness includes a range of activities, including examining our personal relationship to the work, exploring existing policies and processes that dictate how we engage community, and determining how we will value the contributions of partners external to our organization.

We strongly recommend orienting your work along a continuum of involvement, such as The Spectrum of Community Engagement to Ownership, because it encourages organizations to start where they are and build over time.[8] **There is no one right way to work toward increasing community involvement in decision-making. Our guidance is simple—get on the continuum, and work to move right.**

The Spectrum of Community Engagement to Ownership

▶▶▶ INCREASED EFFICIENCY IN DECISION-MAKING AND SOLUTIONS IMPLEMENTATION ▶▶▶ EQUITY

STANCE TOWARDS COMMUNITY	IGNORE	INFORM	CONSULT	INVOLVE	COLLABORATE	DEFER TO
	0	1	2	3	4	5
IMPACT	Marginalization	Preparation or Placation	Limited Voice or Tokenization	Voice	Delegated Power	Community Ownership

Used with permission from Facilitating Power

> If you're toward the left side of the spectrum and committed to doing more, start by assessing your current capacity with the resources in What's Next? This section has tools to help you carefully consider the community you are intending to engage and your institution's history of engaging with them. Considerations for next steps are different depending on the group; for example, see Working with Tribal and Indigenous Data.

---

[8] Gonzalez, R. (2020). *The Spectrum of Community Engagement to Ownership.* Facilitating Power.

Building Social License

As we move to the right on the spectrum of community engagement, we are building trust and what is sometimes called social license, or social approval, to operate data infrastructure. People support what they help create, and in an ideal context, a data integration effort would have social license from the community at large and informed consent from individuals to use data to drive decision-making. However, there are significant barriers to this ideal state. We strongly encourage a developmental approach—start small and grow.

Social license comes from an effort's perceived legitimacy, credibility, compliance with legal and privacy rules, and overall public trust. Earning it requires dedicating time and resources to develop relationships, source and incorporate feedback, and engage with diverse partners on an ongoing basis.

So how do we build social license? How do we move from **ignore** to **inform** to **consult** to **involve** to **collaborate**? The answer, in short, is participatory data governance.

Working Toward Participatory Governance

The core work of centering racial equity throughout data integration is participatory data governance—deliberate conversations and shared decision-making about data use with folks who have diverse experiences, expertise, and skills. This is a simple and intuitive concept, but incredibly challenging to operationalize and implement.

> Data governance is the people, policies, and procedures that determine how data are managed, used, and protected.

Participatory data governance is based upon community members having a voice in decisions. Governance can be made more participatory and robust when many voices are involved—particularly those represented in the data—and when policies and procedures are set up to allow for public deliberation and power sharing. There are many models of participatory governance, with the most common being structures that are broadly focused on operations (e.g., advisory committees) and others that are more focused on specific projects (e.g., participatory action research). See more models and approaches in Resources and in the Work in Action.

The decisions made through participatory data governance are many and impact how equity and community voice are centered at every stage of the data life cycle.
Examples of these decisions include:

- Who sets meeting agendas?
- How are research questions developed and prioritized?
- Who reviews a data request?
- Who can request access to the data?

- Is there a fee for data access?
- How is data quality evaluated?
- How are metadata documented and shared?
- How are quantitative data contextualized?
- How are findings validated prior to release?

While it can be appealing to make many changes all at once, there is a greater chance of sustained success when you start small and build over time, with consensus and broad participation. The following are four simple questions that can be used right now to guide any prospective use of data, whether you have participatory governance in place or not.

When considering data collection, access, integration, or use, ask:

- **Is this legal?**
- **Is this ethical?**
- **Is this a good idea?**
- **How do we know and who decides?**[9]

Note: Question #1 is the minimum standard and can often be answered by legal counsel. Questions #2 and #3 require conversation and deliberation. No one can (or should) consider these questions in isolation. Question #4 indicates whether data governance is in place and how participatory it is in practice. It is important to note that data governance can and should evolve over time.

Assessing Risk & Benefit

Data collection, access, and use comes with inherent risks and also offers tremendous benefits. One of the primary goals of data governance is to carefully assess the risks and benefits of each particular data use. Ethical and equitable use requires evaluating intended and unintended consequences and working to mitigate identified harms. While this should be done for each particular project or analysis you consider, it can also be useful to think about the benefits, limitations, and risks of administrative data use overall.

Benefits

- **Whole-person, longitudinal view:** Using multiple sources allows a holistic view of individual experiences and outcomes across programs and, potentially, across time.
- **Whole-family view:** Administrative data linkages can enable us to better understand experiences and outcomes across a family or household unit.

---

[9] See Hawn Nelson, A. & Zanti, S. (2023). Four Questions to Guide Decision-Making for Data Sharing and Integration. *International Journal of Population Data Science*.

- **Scale:** Administrative data can allow for a population view, rather than a sample, and is therefore less vulnerable to certain forms of bias, such as nonresponse.
- **Time & cost:** Data reuse is less time- and resource-intensive than collecting new data.

Limitations

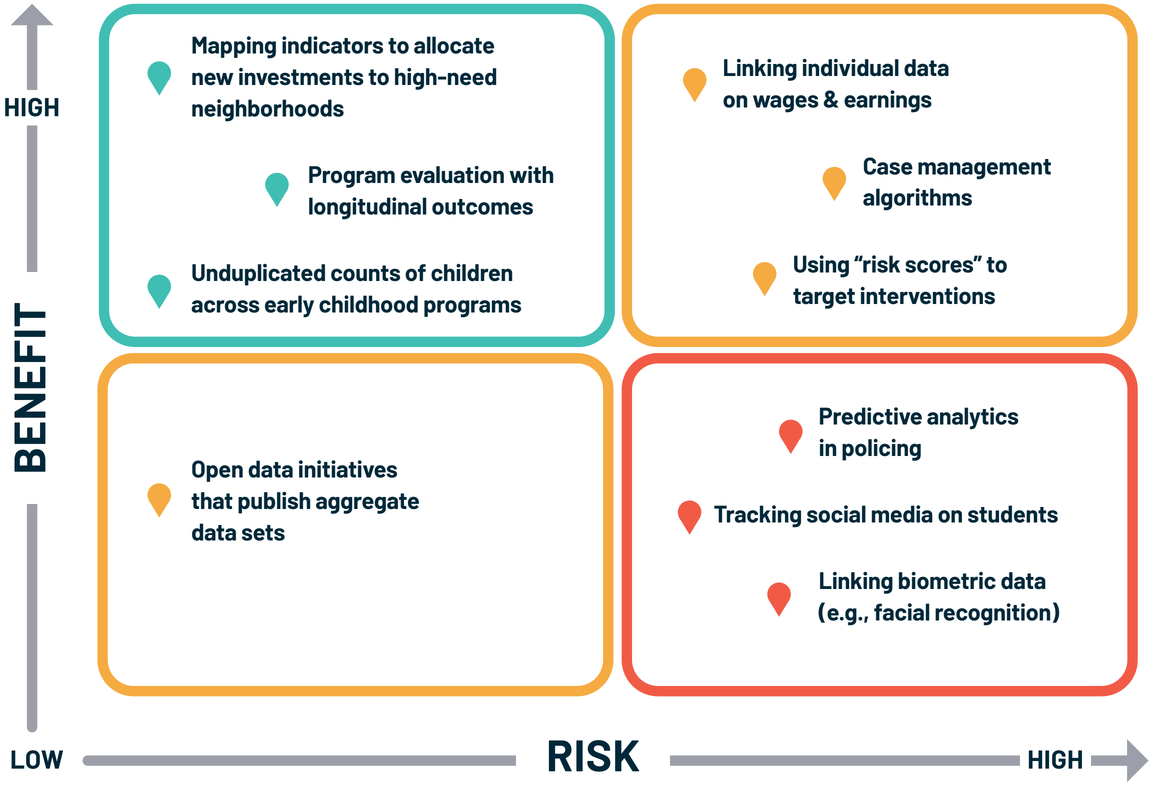
- **Reusability:** Administrative data are collected to meet operational and reporting needs and may not accurately represent the concepts or outcomes an analytic plan aims to measure.
- **Quality:** Data quality issues are common, including missing data and insufficient data documentation, leading to issues of reliability and validity.
- **Depth:** These data are often one-dimensional and may need to be paired with qualitative and other forms of data in order to address deeper questions about causal relationships, client experiences, contextual factors, etc.
- **Access:** Gaining access is often difficult and time-consuming.

Risks

- **Privacy disclosure:** Any transfer of data includes the risk of data being accessed improperly.
- **Misuse of data for research and evaluation:** Without sufficient data documentation, analysts may misuse or misinterpret data.
- **Replicating structural racism:** Administrative data are collected during the administration of programs and services for individuals in need of social supports. These data include people who are disproportionately living in poverty, who, as a result of the historical legacy of race in America, are disproportionately Black, Indigenous, and people of color. Seeing data as race-neutral is inaccurate, and such views could lead to system-level data usage that unintentionally replicates structural racism.
- **Harming individuals:** Certain uses of administrative data carry particularly high risks of causing personal harm. These include uses that provide case workers, service providers, teachers, law enforcement, etc., with personal information that could lead to biased treatments, punitive action, or lengthened system involvement.
- **Harming communities:** Use of administrative data, especially when mapped or otherwise represented spatially, can create or deepen community stigma. When analysts fail to understand and acknowledge the discriminatory practices and structural causes of disparate outcomes by race or geography, they risk using administrative data in ways that perpetuate harmful, deficit-based narratives.

While the particulars are important, we have identified broad categories of use on the **risk vs. benefit matrix** below. For example, projects that involve low risk and high benefit, such as a longitudinal program evaluation, indicator projects, or generating unduplicated counts across programs, are generally a good idea and an easy starting point for collaboration. Conversely, projects that are low benefit and high risk, such as sharing data from menstrual cycle tracking apps, using social media content for predictive policing without the opportunity for public comment, or police surveillance using biometric data often without disclosure, should be considered with the utmost caution, and in some instances data sharing should not proceed.

Examples included in this toolkit are not meant to be used as binary yes/no decision-making tools, but rather as guides for thinking about and talking through intended and unintended consequences of civic data use with a variety of voices and perspectives.



Including more voices as part of your deliberation about the risk vs. benefit of a use case will both strengthen your social license and also require more time and resources. This is a delicate balance that each of us must walk with on our path towards more participatory governance.

We encourage:

- Assessing organizational readiness for community involvement along a spectrum; see What's Next? for a framework and tools to help determine next steps
- Working to build social license developmentally, with participatory governance as the goal
- Using the four questions as a starting point to assess use cases
- Carefully considering risk vs. benefit for data collection, access, and use

WORK IN ACTION:
Foundations for Community Involvement

The Work in Action[10] sites featured in this section have started small and expanded with purpose and intention over many years of sustained effort. Of the seven data integration efforts, five are local, based at the city and county level. One of the local efforts is hosted by a government agency, one by a university, and three by community-based organizations with strong ties to government agencies. The other two efforts operate at the state level, one from within state government and one outside of state government.

Though these sites vary widely in terms of their areas of focus, scale, and staffing models, all have worked slowly and steadily to center racial equity and weigh the risks and benefits of data use. Each of these sites has also taken the time to carefully develop and refine participatory governance structures. They have demonstrated great impact and also made mistakes and course-corrected along the way. While the sites we feature are impressive, what they are doing is not magic, but rather hard and committed work. We hope these examples of Work in Action both inspire your efforts and provide proof that these hard things are doable—even in politically complex environments.

Each summary is told from the perspective of the people doing the work. They provide a description of their work moving toward the right on the Spectrum of Community Engagement, as well as lessons learned. You can also access long-form Work in Action describing participatory governance approaches here. In the next section, we will work stage by stage through the data life cycle to help you on your journey, with shorter and more targeted examples of how sites are centering racial equity throughout the data life cycle.

Baltimore City Youth Data Hub

Bridget Blount & James Sadler

The Baltimore City Youth Data Hub integrates individual-level data between Baltimore's Promise (a collective impact nonprofit), Baltimore City schools, and the City of Baltimore. Authorized by state legislation in 2022, our approach at the Youth Data Hub brings together communities, providers, policy makers, researchers, and young people to inform and make decisions about the creation and implementation of programs and policies designed to eliminate disparities and drive equitable outcomes for all of Baltimore's youth and families.

With community input, key partners began working on the Youth Data Hub in 2015. We collaborated to develop a common mission, establish protocols for data sharing, and build an integrated data system that can provide robust, timely, action-oriented information to benefit Baltimore.

Our early, foundational work fueled landmark progress over the past two years, with the help of the AISP Equity in Practice Learning Community. The Youth Data Hub focused on building a strong governance structure, ensuring substantial youth engagement in the work, and digging into data sharing projects to serve our young people better. A highlight of our governance structure is our Community Research and Action Committee (C-RAC), an intergenerational body made up of 14 residents, half of whom are between the ages of 16 and 26. C-RAC members, who are compensated

---

[10] We use the term Work in Action, rather than exemplar, to feature site-based work that is ongoing and rooted in learning and humility, acknowledging that all work at the intersection of data infrastructure and racial equity is complicated and contextual and not necessarily replicable.

monthly for their work, enrich and drive the Youth Data Hub forward by providing the context of their own lived experiences, facilitating community interpretations of findings, and ensuring results are accessible to multiple audiences.

Lessons Learned:

- The participation of youth and community members is a necessity to ensure equity, accountability, historical context, and vision.
- Building relationships, trust, and buy-in among partner institutions and community stakeholders is vital for sustainability and building an equitable data infrastructure.
- Stay focused on the desired outcomes rather than being tied to a single approach, especially when navigating legal challenges; be adaptive but remain committed to the overall mission.

Data Collaborative Serving Children & Youth

In a large county in the southeastern U.S., partners have fostered an ecosystem where local government can bring system participants and system professionals together as co-researchers to improve child-serving system processes and outcomes. Since 2018, the partnership has completed three community participatory action research (CPAR) projects with youth, parents, and system professionals in child welfare, juvenile justice, and behavioral health. CPAR projects have resulted in policy, system, and narrative change by centering the expertise of system participants and by building relationships that nurture trust and share power across different interest-holding groups. They have embraced asset-based approaches, offered stipends for participants to acknowledge nontraditional expertise, and designed experiences that bridge social divides and differences in order to create a culture of possibilities. The CPAR projects also inspired a local business chamber to take a similar approach, partnering with community members as co-researchers in their own work.

Elements vital to success: Dedicated and visionary collaborative staff; supportive collaborative leadership; strong relationships with program staff and system partner organizations.

Barriers: Resistance to shift from power and resource hoarding to power and resource sharing with community; limitations of engagement due to state policies around language.

What we wish we had been told:

- You will need to expand your frameworks around participatory rigor, rather than methodological rigor.
- You will have to build the organization's capacity to support payment for community engagement— specifically, new ways of approaching procurement, contract design, and renewals.
- Your organizational staff will need support to engage in authentic, trauma-responsive relationships and conversations.

Lessons learned: Co-researching can promote healing, learning, and more innovative system solutions.

Charlotte Regional Data Trust

The Charlotte Regional Data Trust is a partnership that links administrative data across service and organizational silos to provide information the community can act upon. The Data Trust was founded in 2004 as a 501(c)3 nonprofit with the mandate to maintain the confidentiality and useability of these important data. Today, the Data Trust maintains this commitment and is governed by a board of directors. The Data Trust allows people to request access to the data and employs a participatory governance model to review and make decisions about data requests. Since its founding, this process has been carried out by the Data and Research Oversight Committee (DAROC), a subcommittee of the Board of Directors. Members of DAROC include subject matter and methodological experts as well as staff of data-contributing agencies, which allows agencies to retain oversight of the data.

In 2022, we began developing a Community Data Advisory Committee (CDAC). The vision of the CDAC is to create intentional and formal processes to share oversight power over Data Trust data with people whose experiences are represented in the data. This practice integrates community expertise and insight into more stages of the data life cycle, helping to ensure ethical use, accurate interpretation, and actionable dissemination.

Participatory governance is the foundation of the Data Trust. Governance activities are staffed by local affiliates, and governance committees are working subcommittees of the Board of Directors. Chairs of each governance committee (DAROC, CDAC, and ad hoc committees) are also part of the executive committee of the Board so that committee and board leadership are informed of ongoing work, fostering strong, reciprocal communication and collaboration.

By design, the Data Trust is deeply embedded in the community in myriad ways—working with partners on data and research projects, participating on committees, serving as a data resource for the community, and establishing data sharing partnerships. All of these efforts contribute to participatory and community-engaged governance that ensures that the Charlotte community guides the work of this valuable community resource.

Lessons learned:

- Lead with the why, not the what
- Lead with the why, not the what.
- Start small with people willing to work through the "messiness."
- It doesn't have to be hard to find great leaders. Work with community members to identify people who already have trust.
- Be prepared to take a close look at the data in your data system, and ask new questions: Who is represented and overrepresented? Who is missing? Why? And remember that this is all just a starting point for deeper discussion and understanding.
- Contingency planning is essential to ensure that the governance work is sustained.
- Don't do it alone. Find community both in and outside the work.

State of Connecticut

Office of Policy and Management (OPM), Department of Social Services (DSS), and Office of Health Strategy (OHS)

Scott Gaul, Katie Breslin, & Pauline Zaldonis

In June 2021, Connecticut declared racism a public health crisis and put in place initiatives to decrease disparities and improve demographic data collection with strong community input. Since 2022, the state of Connecticut's Office of Policy and Management (OPM), Department of Social Services (DSS), and Office of Health Strategy (OHS) have been working as part of the Equity in Practice Learning Community (EiPLC) to build a foundation for securely and responsibly sharing administrative data to address the crisis, as well as to work on state readiness for broad community engagement. Members of our EiPLC team have focused on existing collaborative data efforts, such as the P20 WIN data system and the All Payers Claims Database. Together, they have worked to develop standardized methods of collecting Race, Ethnicity, and Language data, developed and implemented a Data Protection, Privacy, + Equity Impact Assessment tool, and co-created and disseminated new guidance documents, such as the Equity Awareness in Data Reporting and Data Visualization and Accessibility Guidelines, across state agencies. OPM believes that all of these efforts have built their readiness to institute a Resident Advisory Board, composed of current and former state service recipients, to build accountability and transparency into the data sharing process for those overrepresented in the data. CT hopes the board will also be a mechanism for building lasting relationships and empowering service recipients with more information and tools to support advocacy. Learn more here.

Elements vital to success and sustainability:

- Strong interagency relationships (through EiPLC work and other interagency efforts)
- Approach the work steadily and consistently (having steady habits has served us well)
- Normatively tying an equity lens into other State-wide data activities and practices (e.g., State Data Plan, Agency Data Officer convenings and community of practice)
- Strong commitment from EiPLC group members
- EiPLC group members working on multiple threads in parallel, driving different streams of action, leading to wider impact

Barriers:

- Limitations in ability to implement some changes as state agencies
- Ability to develop guidance and resources but challenges in enforcing these guidelines
- Lack of funding

Lessons learned:

- Interest is not the same as buy-in.
- Collaboration is key because this work cannot be done in isolation.

Data You Can Use (DYCU), Milwaukee, Wisconsin

Victor Amaya, Executive Director

Data You Can Use (DYCU) is an independent nonprofit in Milwaukee, Wisconsin, dedicated to advancing community access to the data and skills needed to drive informed and equitable decision-making. DYCU empowers local nonprofits, funders, and residents by making data understandable and actionable to improve neighborhood and community conditions. DYCU services include neighborhood portraits, data chats, program evaluations, and training to support community-driven solutions. Each is unique in its purpose, but all are designed to grow participant capacity for collaborative data use. For example, data chats—small community conversations about data designed to draw out resident perspectives and interpretations—were designed by DYCU to improve two-way communication between researchers and community members. With a focus on blending data and resident expertise, DYCU provides services to bring nonprofits and community members in structured conversations and engagement around their data, the meaning behind it, and possibilities for use.

Elements vital to success: Training and supporting staff and community members—the more capacity they have the more effective we can be.

Lessons learned:

- There is no data without stories and no stories without data. Combining data with resident stories provided a more comprehensive understanding of community challenges.
- Establishing clear expectations and structure to resident participation ensures that projects are relevant and impactful.

What we wish we had been told: Securing diverse revenue streams is crucial for long-term stability.

The Hartford Data Collaborative (HDC)

Kate Eikel & Michelle Riordan-Nold

The Hartford Data Collaborative (HDC), staffed by the Connecticut Data Collaborative (CTData), facilitates data sharing and integration among a network of Hartford area nonprofit organizations, government agencies, and philanthropic partners. HDC was founded in 2019 to connect data across programs serving Hartford youth to provide a more holistic view of individuals, families, and households in Hartford. In 2022, HDC launched the Hartford Youth Data Fellowship program, which recruited youth from across the HDC network and from other trusted partners. Through the fellowship, youth leaders serve as co-researchers, developing community data projects and receiving data training.

The original intention of the Hartford Youth Data Fellowship was to incorporate community voice into data integration projects; however, we realized that the training was also opening up new college and career options for the youth fellows. With this in mind, we adapted the fellowship to include guest speakers and career development resources to support the fellows as they move into adulthood. Over the course of the first three cohorts, HDC has developed a framework for youth co-researchers to drive the research agenda while building skills that open up new pathways for them going forward.

What we wish we had been told:
Youth (or community) participation can benefit the participants as well as the data effort. When participatory work is mutually beneficial, there is more buy-in from those community members.

Lessons Learned: Youth fellows bring in great ideas, but often have moved on to other opportunities by the time the data can be shared and integrated. HDC is working on a new project that will allow fellows to engage with integrated data earlier in their tenure.

Metriarch, Oklahoma

Jacqueline Blocker

Metriarch® is a nonpartisan research and policy organization shifting the narratives that drive conversations about women's health in Oklahoma. This organization is part of the community-to state-level solution to a domain understanding gap around women's public health. Metriarch is unique in that it not only focuses on women's bodily health, but the social determinants and drivers of public health as well. Layers of data and legislative analysis impact smart policy. Metriarch brings together statewide decision makers to ensure that policies are robust and effective, and properly address the needs of our respective communities.

Metriarch's mission is to shift narratives about women's health through mixed methods data analysis and tracking synthesis, policy evaluation, and outreach events. The organization's diverse network of partners and friends facilitates and encourages statewide collaboration on public health issues lending itself to a more holistic approach to evidence-based solutions and policies. Metriarch is incubated within Take Control Initiative, LLC (TCI), an organization focused on removing barriers to access to contraceptives in Tulsa, Oklahoma. TCI was established over a decade ago to address rising teen pregnancy rates. TCI and other community partners quickly recognized the need for accurate data and research regarding women's health, yet the organizations did not have the bandwidth to focus on data curation and analysis. More importantly, TCI and its partners realized it is difficult to find data related to women's health and sometimes even harder to understand it.

Metriarch was born out of a need for a data intermediary to democratize publicly available data regarding women's health and launched in 2019 with an interactive event entitled Lady Charts, the data jam in women's health. This bi-annual event brings organizations and individuals from around the state to learn from experts about issues such as substance use disorder and pregnancy and paid family leave. Metriarch hosted its third Lady Charts in October 2023 with over 200 attendees.

Metriarch is a data haven for women's health in Oklahoma. Visit metriarchok.org for more information.

Resources: Foundations for Community Involvement

Digital Defense Playbook: Community Power Tools for Reclaiming Data, 2019, Our Data Bodies

Engagement, Governance, Access, and Protection (EGAP) Framework, 2021, The Black Health Equity Working Group

A Path to Social Licence: Guidelines for Trusted Data Use, 2017, Data Futures Partnership

Participatory Data Stewardship, 2020, The Ada Lovelace Institute

A Roadmap for Effective Community Engagement in Healthcare, 2024, INSPIRE, Camden Coalition

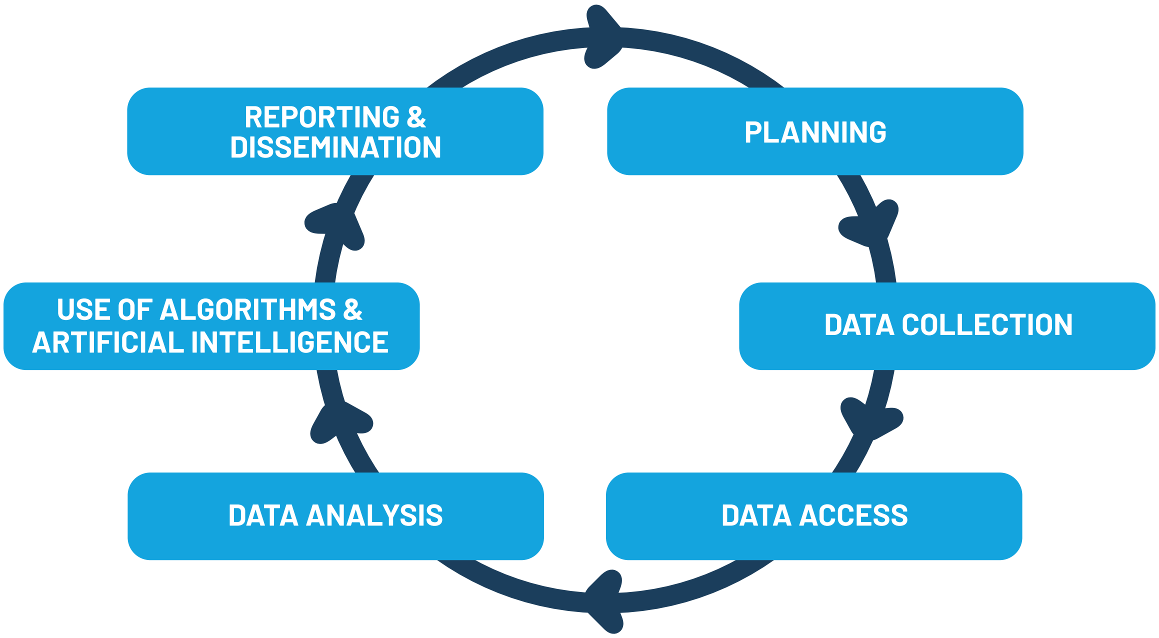
The Spectrum of Community Engagement to Ownership, 2020, Facilitating Power

The Trust Toolkit: Spitfire's Toolkit for Civil Society Organizations to Earn and Repair Trust, 2024, Spitfire Strategies

Youth Collaboration Toolkit, 2019, True Colors United

Centering Racial Equity Throughout the Data Life Cycle

Centering racial equity throughout data integration is not a single, discrete step, but rather an ongoing process at each stage of the data life cycle—planning, data collection, data access, data analysis, use of algorithms and artificial intelligence, and reporting and dissemination. Each stage presents new opportunities to bring a racial equity frame to data integration, as well as new challenges and considerations. Depending on your role, you may have more experience (and leverage) in some stages than others. We encourage you to focus on the pieces most relevant to your work and to consider allies and partners who have the potential to shift practice where you do not.



The following sections provide an overview of racial equity considerations throughout the data life cycle, examples of positive and problematic practices, and brief examples of Work in Action. The Work in Action highlight current examples of how organizations from across the AISP Network and beyond are centering equity at each stage of the data life cycle.

Racial Equity in Planning

Planning is the first stage of the data life cycle and includes all the work to prepare for future stages. While this includes project planning and scoping, it also includes preparations even further upstream, such as assessing organizational readiness (see What's Next?); identifying partners; articulating a purpose for data integration; identifying relevant legal authority (see Finding a Way Forward) and considering duty of care;[11] developing understanding of the local racial, social, and

---

[11] Duty of care is a legal obligation to act with "reasonable" care to prevent foreseeable harm to others. For data sharing and integration, the duty of care involves providing a "reasonable" amount of protection (e.g., safeguarding data assets) to constituents, at a minimum, but it is also implicitly owed to the community we aim to serve.

historical context; and creating processes to evaluate risk vs. benefit for each project or data use. It is a common mistake to let the forward momentum and looming deadlines at the outset of a data project overshadow upfront (often internal) equity work. **Incorporating a racial equity lens during the planning stage sets the foundation for embedding racial equity throughout the data life cycle, making it a critical first step.** Below, we've broken out positive and problematic practices into what we see as three core components of planning.

Individual and Organizational Readiness refers to the work of individuals and teams within institutions to operationalize values, ensuring alignment between principles undergirding the work and operations (see Baltimore's Promise and the Allied Media Project in the Work in Action section below). This also involves individuals and the broader team reckoning with their identity and biases, to develop and refine skills needed to engage in equity and engagement work authentically, without inflicting harm on community members (see Open Data Charter). Readiness work also includes efforts to ready community members for active roles in data projects (see Black Researchers Collective).

Team Formation and Governance is about getting the right people and processes in place to foster equitable data use (see I2D2 and Mecklenburg County Community Support Services). This includes defining both a shared purpose and individual roles, building trust in relationships, ensuring that power dynamics are illuminated, and establishing transparent decision-making processes. Building strong governance and staffing upfront will ultimately support more equitable and participatory decision-making when important questions arise throughout the data life cycle. For example, many data sharing efforts wrestle with whether individual-level consent for data sharing and use is required when legal requirements are unclear. Such challenges can be minimized through the development of strong, collaboratively generated governance agreements that clearly lay out processes for decision-making.

Lastly, **Project Planning** with an equity lens involves learning about the context and history surrounding data and systemic racism (see Embrace Boston and Connecticut Office of Early Childhood). It then requires that we align our research questions and approach to community needs, the available resources and data, and areas where there is traction for change (see Philadelphia Monument Lab). As shown in the Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems, effective planning can also involve highlighting the risks of data sharing along with the potential benefits. See the Companion Workbook (1.III) for guidance on facilitating this type of discussion. Thoughtful planning that centers equity from the start paves the way for more responsible and impactful data use throughout the rest of the data life cycle.

Positive and Problematic Practices: Planning

POSITIVE PRACTICE	PROBLEMATIC PRACTICE
Individual and Organizational Readiness	
Engaging in individual education and self-reflection around personal identities, implicit bias, racism, power, and oppression.	Assuming that data are neutral, that individual identity does not influence data use, and that analyses can be conducted free of bias.
Ensuring alignment between mission, vision, guiding principles, and organizational policies around employee benefits and well-being.	Not acknowledging the labor required in this field and encouraging burnout practices.
Developing organizational understanding of sociological perspectives along with strategies to center racial equity.	Moving projects through the data life cycle before understanding the broader social, historical, and political context of data access and use and the organization's place in the field.
Building relationships between community, administrators, researchers, and technologists over time so that they are prepared to work together collaboratively and effectively.	Using deadlines or grant deliverables as an excuse to avoid investing in relationships.
Preparing community members and formally trained researchers to take an active role in data projects and providing opportunities for all participants to speak, listen, and build knowledge.	Failing to demystify data and maintaining the false idea that working with data is only for those with specialized training.
Team Formation and Governance	
Creating participatory data governance that intentionally involves a diversity of perspectives and skillsets—community members, subject matter experts, agency staff, methodologists, etc.	Using only token "representation" in data governance processes.
Setting clear mission, vision, and guiding principles to ground data governance processes and formalize the role of community voice, oversight, and ownership in decision-making.	Refusing to cede power and be accountable to the community whose data are being used.
Staffing the data effort with people who reflect the population or jurisdiction the data represent.	Not compensating people who make meaningful contributions of their lived experience,* knowledge, time, or skills.

* It is important to note that we all have lived experiences, and the term "lived experience" should be relevant to the topic of study and not used as coded language to imply marginalized identity.

Using a framework to clearly identify and articulate roles and power (e.g., RAPID is a decision-making tool to determine roles with the power to Recommend, Agree, Perform, provide Input, and Decide throughout the life of the project).	Not being clear upfront about who holds the power to frame research questions, make important decisions, and determine what types of policy/practice responses are "on the table" (or not).
Ensuring that data governance processes are accessible, inclusive, and engaging.	Creating an environment that is not attuned to the current level of knowledge, accessibility needs, or interests of the group engaging with data.
Taking time to recognize the unique knowledge and talents project team members—especially community participants—bring to the work.	Giving disproportionate power and authority to certain position titles or academic credentials.
Project Planning	
Researching, understanding, and disseminating the history of local policies, systems, and structures relevant to the effort, including past harms and their potential to be replicated.	Using only administrative data to describe the problem, without including historical and contextual information that supports "multiple ways of knowing."
Creating space during the planning process to envision a future state that upholds shared ideals.	Jumping to analysis without thoughtful deliberation about the future state that the project will contribute to.
Developing research questions that address needs identified by community partners and framing the questions in ways that do not reinforce harmful narratives.	Relying on academic/institutional partners to frame the problem and research questions while failing to engage community partners.
Connecting research questions to a clear plan of action to improve policy, services, and outcomes.	Resourcing projects that are not aligned with areas of traction for real policy/practice change.
Coming to general consensus among partners on the approach, methods, metrics, key resources/literature, etc.	Requiring conformity rather than consensus in order to move forward.
Managing expectations around the available data, what questions can realistically be addressed, and how long it may take to reach desired actions and outcomes.	Only considering the effort a success if "big" outcomes are achieved, rather than seeing the inherent value in engaging in a process of authentic collaboration, listening, and responding in real time.
Supporting shifts to the funder ecosystem to enable shared decision-making and collaborative funding.	Pursuing grant/philanthropic funding for projects that do not address a community priority or need.
Using the continuum of engagement to reflect on the project and ensure that processes honor voice and agency.	Not taking time to debrief and reflect on a project with community members involved to hear what they would do differently next time.

WORK IN ACTION:
Planning

Individual and Organizational Readiness

Baltimore's Promise, Operationalizing Organizational Values

Bridget Blount & James Sadler

At Baltimore's Promise, a citywide collaborative that uses data to improve outcomes for children, salaries are determined by the responsibilities and level of authority for each position, as opposed to being negotiated around each individual candidates' qualifications or pedigree. This standard, transparent compensation structure aims "to mitigate pay inequities that are in part driven by systemic and implicit racism and gender bias impacting negotiation and career trajectories over time." Further, Baltimore's Promise provides benefits designed to honor the humanity of staff members. Since 2020, the organization has provided scheduled paid "BP Holidays" in addition to existing paid holidays and individual paid time off. BP Holidays are intended to "encourage our hard-working team to rest and take time for themselves." This commitment to both team- and individual-level care is also evident in their guiding organizational values, which balance the importance of responsibility to the team and a commitment to the hustle necessary to the mission.

Allied Media Project, Default to Private Browsing

Toni Moceri, Molly Leebove, & Chiara Galimberti

The Allied Media Project website defaults to private browsing, which is consistent with the tremendous value placed on collective care and responsibility in their mission, vision, and guiding principles. Unlike the case with most websites, visitors do not have to change their settings before visiting to protect their data; instead, users get the opportunity to opt in to help improve the website through data sharing. The "Private Browsing" widget appears in the bottom corner of the page with a clear explanation of what it means to enable data sharing and the exact data that will be collected. This feature allows all users of the website to engage critically and with consent. The page also recommends ways users can protect their information when browsing other websites.

Open Data Charter, Mission & Vision Statement (ODC)

Cat Cortes

The Open Data Charter (ODC) works with government agencies across the world to help with their "openness journeys," encouraging a shift toward governments being "open by default." Their mission emphasizes their dedication to both government transparency and individual privacy rights. They acknowledge that public sector data exist on a spectrum from fully open to protected, and help government partners figure out how to "publish with a purpose" when it comes to open data. Governments endorsing and adopting the ODC principles acknowledge they are working toward more open access to data with strong governance and accountability. This is a strong foundation on which to build public engagement about the sharing and integration of more restricted data as well. Read their entire mission and vision statement here.

Black Researchers Collective, Community Capacity Building

Glenance Green & Shari Runner

Black Researchers Collective is a capacity-building collective “taking research to the streets.” They equip communities with tools to understand research, data, and policy, so these communities can be their own self-advocates. Offerings like the *Research in the Streets* podcast, customized organizational trainings, coffee chats, and community workshop intensives aim to teach listeners and participants how to conduct their own research, harness their power, and tell their own stories. The Collective empowers people with confidence and skills to be civically engaged, and their work offers lessons for agencies seeking to build data-driven community conversations and bring new voices into their work.

Team Formation and Governance

Iowa’s Integrated Data System for Decision-Making (I2D2), Collaborative Inquiry Agenda Setting

Heather Rouse

Iowa’s Integrated Data System for Decision-Making (I2D2) integrates human service, education, and health data to identify priority issue areas and inform program and policy support. I2D2 is led by a Governing Board composed of directors from the agencies that contribute data: the Departments of Education, Health and Human Services, Workforce Development, and Economic Development Authority, and the Iowa Head Start Association. The Board and their co-chairs—the office administrator from the Early Childhood Iowa initiative and the Director of the Integrated Data System Resource Center (IDRC)—determine meeting agendas and review recommendations from the IDRC staff team and the Data Stewardship Committee (DSC). The DSC is appointed by the Governing Board, and is composed of subject matter experts from each department who know the data (and the people and programs in the data) well. DSC members support operational policies, as well as engage in cyclical co-creation of an inquiry agenda to drive the system. These leaders bring their departmental priorities to discussions, support developing project ideas that meet the departmental needs, and also work to ensure that analytic findings from I2D2 are translated into practice by data-contributing agencies and other partners. This often means that, at the end of an analytic cycle, new questions are generated and new data integrations or analyses are initiated to facilitate deeper understanding of population and program needs. The tiered I2D2 governance structure institutionalizes the input of those contributing data into the system, and gives subject matter experts agency to drive inquiry and better support their communities. Learn more about I2D2 at their website and on their AISP Network page.

Mecklenburg County Community Support Services, Building Staff Capacity

Mary Ann Priester

In 2014, Mecklenburg County (NC) Community Support Services invested in two new positions tasked with improving data collection, access, and community use of data: the Homelessness Management Information System (HMIS) Administrator and the Homelessness & Housing Data and Research Coordinator. Both positions demonstrate an investment in the organization’s commitment to collecting high-quality, useful, and timely data, and putting those assets into the hands of community organizations and members. The HMIS Administrator works to expand community access to high-quality, accurate data on housing and homelessness by implementing best practices for data collection and analysis, while also ensuring that high-quality data are available to partners. The Homelessness & Housing Research Coordinator connects constituents with research and data to inform programming, policy, and funding decisions related to housing instability and homelessness. In 2018, as a result of expanded community use of HMIS and changes to the Department of Housing and Urban Development's expectations related to HMIS data, Mecklenburg County, as the HMIS Lead, applied through the Continuum of Care grant program to fund one additional staff person. During this time, Mecklenburg County Community Support Services also repurposed an existing internal role to create a third HMIS staff position. Both positions were filled in 2019, bringing the total number of HMIS staff to four. While they work collaboratively, each has designated roles that focus on the following areas of expertise: (1) training and data collection, (2) reporting, and (3) privacy and governance. Further, the Homelessness & Housing Research Coordinator has implemented a participatory approach and partners with community stakeholders, data partners, and persons with lived experience of housing insecurity to expand and enhance the communication and use of homelessness and housing data in Mecklenburg County.

Learn more about Community Support Services use of data to end homelessness, including the Housing & Homelessness Dashboard, and read more about their partnership with the Charlotte Regional Data Trust.

Project Planning

Connecticut Office of Early Childhood (OEC), Statement on Racial Equity

Chanae Russell & Elena Trueworthy

In 2020, the Connecticut Office of Early Childhood (OEC) created an agency committee on Advancing Equity & Anti-Racism. Multiple subcommittees were also formed, including a Communications subcommittee that was tasked with creating an OEC Statement on Racial Equity in 2021. The statement addresses how systemic racism creates barriers for marginalized communities to access OEC resources. It also clearly states that anti-racism and equity are integral to the healthy development of all children, driving the agency's commitment to remove barriers through responsive communication with constituents about what is and isn't working. The agency is committed to living this statement in practice and references it often during external work, including strategic plans, contracts, and project action steps. OEC often refers to the Racial Equity Statement internally, sharing it with new hires and revisiting it during data and research projects to ensure its continued application. Read the full statement here.

Embrace Boston, Harm Report

Jenny LaFleur

Embrace Boston's Harm Report explores how centuries of systemic racism impact the contemporary experience of Black Bostonians. By looking across seven "injury areas"—including culture and symbols, health, education, and others—the report makes the history and harms of the past visible in the present, and argues that symbolic and material reparations are required to pave the way toward a better future. The report is a strong example of the type of historical exploration that can build sociological understanding and spur self-reflection for public agencies and other institutions as they ready themselves to engage community members without doing more harm. In their words: "Let's ensure that we are on the right side of history. May the work of repair and the imagining of a better future begin." Learn more and read the report here.

Federal Agencies, Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems

In April 2023, executives from five federal agencies released a joint statement on the potential harms of automated systems and artificial intelligence. The agencies represented operate at the federal level and are involved in efforts to “protect civil rights, fair competition, consumer protection, and equal opportunity,” including in response to emerging technology. Their joint statement acknowledges the ways automated systems can perpetuate existing harms or generate new forms of discrimination, and clarifies the role laws can play in mitigating harm. It also recognizes the value of innovation without minimizing the real risks that come with emerging technologies. This unified statement is a powerful demonstration of interagency collaboration and thoughtful discernment and articulation of the risks and benefits of using this technology. To have this many partners speak out against the neutrality of technology and commit to proactive enforcement of laws sends a powerful message to those producing new tools and products. Read the full statement here.

Philadelphia Monument Lab

Paul Farber

In 2017, Monument Lab—a Philly-based nonprofit public art, history, and design studio—launched a participatory research exhibition with artists and researchers to ask: “What is an appropriate monument for the current city of Philadelphia?” The goals of the project were to engage public artists in the creation of a series of monument prototypes for display across the city and to invite Philadelphians and visitors to respond to the central question with their own handwritten monument proposals. They broke down their planning into five key components: what question is asked, who asks the question, where the question is asked, how the question is asked, and what is done with the answer. They partnered with Mural Arts Philadelphia, as well as 10 other city agencies and institutions, to produce the citywide exhibition driven by this artistic research project. At the core of the work was a desire for every response to be treated as city data, worthy of consideration, and a piece of the city’s history. Community engagement was conducted by trained researchers face-to-face, in public places, and led by people knowledgeable about Philadelphia (including high school and college students, as well as local artists and public historians). Every response was handwritten or drawn, physically handed over, displayed, cataloged, coded, and archived. Ultimately, 4,500 monuments were proposed, all of which are openly available for public viewing. The process culminated in a Report to the City, shared with municipal leaders, in each public library branch across the city, and beyond. Learn more about the project here and Monument Lab’s ongoing work here.

Resources: Planning

What’s Next?

Community First Toolkit Start Guide, 2024, High Line Network

AISP Network Survey Brief: Governance, 2025, Actionable Intelligence for Social Policy

The Data Equity Framework, 2021, We All Count

(Divorcing) White Supremacy Culture, 2021, Tema Okun

Why Am I Always Being Researched?, 2019, Chicago Beyond

Racial Equity in Data Collection

Data collection is the process of gathering information in an organized way. This process can involve collecting primary data or using secondary information from another source. Both primary and secondary data collection have significant benefits and limitations.

Primary data collection requires administering an instrument, such as a form or a survey, to a specific population. This means that the data collection can be designed for the particular needs of the project, ideally with the focus population in mind (see Oregon Health Authority in the Work in Action). Primary data collection for specific populations can also be challenging as a result of cultural norms, stigma, distrust, and fear of misuse that can lead to inadequate response rates and incomplete responses. **Secondary data collection** involves using data originally collected for a different purpose. The reuse of administrative data—data collected during the routine process of administering programs—is commonplace. However, because these data are not necessarily collected for reuse, there are benefits and risks that should be carefully considered (see Assessing Risk & Benefit). Data minimization is an important principle for ethical data use, as collecting, storing, using, and retaining data has implications for both privacy and environmental impact.[12]

Data minimization: The principle of limiting or minimizing the collection, storage, and disclosure of data to only what is necessary to accomplish a specific use.

All data are vulnerable to biases, inaccuracies, and missingness. Bias within administrative data is commonplace and most often takes the form of selection bias (i.e., the individuals included in the data are not random or do not represent the intended population), as these data tend to include communities that are over-surveilled by government agencies. Confirmation bias (i.e., data used to confirm pre-existing beliefs) is also a concern, due to the impacts of unexamined individual, institutional, and systemic racism on data collection—for example, an intake form that does not list a significant racial or ethnic group within the population. Missing, poor quality, or inaccurate data on demographics, including race, ethnicity, language, and disability (RELD) or sexual orientation, gender identity, and expression (SOGIE), can also erode validity and community relevance of the study outcomes.

Administrative data are often collected from intake paperwork, self-reported online applications, service records, and participant surveys, so the information is in different formats and inconsistently defined across agencies, programs, and services. Sufficient metadata (i.e., data about data) is essential to design valid and reliable analytic plans and to harmonize data prior to integration, yet is not often created as part of data collection. This lack of data documentation is a significant risk that can lead to misuse.

At the same time, administrative data also creates opportunities to better capture and understand the experiences of individuals and subgroups. The sheer volume of information available across administrative data sources may allow for programs with low-quality or missing demographic data to utilize higher-quality sources. Integrated administrative data may also enable the exploration of intersectional experiences (e.g., how queer youth of color experience health care) in a way that is not

---

[12] Down on the server farm. (2008). *The Economist*.

possible using a single source. Better understanding of these experiences then allows for specialized policy and program interventions and supports.

Intersectionality refers to the interconnected identities of an individual—including race, gender identity, nationality, sexual orientation, and disability. Collecting demographic data to better understand disparate impact based upon intersectionality can be an important step toward equity.

> *"Intersectionality is simply a prism to see the interactive effects of various forms of discrimination and disempowerment. It looks at the way that racism, many times, interacts with patriarchy, heterosexism, classism, xenophobia—seeing that the overlapping vulnerabilities created by these systems actually create specific kinds of challenges.*
>
> —Kimberlé Williams Crenshaw[13]

An equity lens considers these advantages and vulnerabilities in the data collection and reuse process and how they can be weighed or contextualized appropriately in response. The Work in Action examples below highlight different strategies for centering equity when designing and implementing primary data collection. These strategies range from oversampling populations that are often overlooked because they are small in numbers (see Albuquerque Area Southwest Tribal Epidemiology Center), to enlisting community partners to redesign survey instruments (see Oregon), to changing the standards on what is collected (see Allegheny County and Connecticut), to digitizing and cataloging records to provide public access (see Mapping Racist Covenants). The practices below highlight different strategies for collecting data with equity centered.

Positive and Problematic Practices: Data Collection

POSITIVE PRACTICE	PROBLEMATIC PRACTICE
Designing flexible data systems that can support shifts in data collection and data management needs over time, and consulting the governance structure as updates are made.	Changing data collection practices without documenting and considering the impact on longitudinal analyses.
Protecting data and ensuring data are made accessible for approved uses by creating clear contracts and agreements with technology providers.	Storing data in systems where cost (financial, technical, contractual, or otherwise) impedes the ability to adapt data collection practices, access the data, or migrate data to other platforms (e.g., a tech provider charging for an export of data).
Co-creating a data development agenda with community partners that identifies what data are meaningful and will be collected and reused, and for what purpose.	Pursuing data sources that reinforce deficit narratives, stereotypes, or biases rather than informing meaningful policy and practice change.

---

[13] Quoted in Guobadia, O. (2018). Kimberlé Crenshaw and Lady Phyll Talk Intersectionality, Solidarity, and Self-Care. *Them*.

Involving community members in primary data collection activities where possible and appropriate, and compensating them for their time.	Failing to support and learn alongside community members so that they are empowered to inform data collection and research activities to their full extent.
Prioritizing data minimization.	Collecting data for the sake of having more data, or collecting data that is not sufficiently granular to meet community and project needs (see RELD & SOGIE Data Standards Framework).
Coordinating across institutions and programs to prevent duplicative data collection and use existing data sources where possible.	Being extractive by collecting data in ways that benefit the institution or research team without demonstrated benefit to the community (e.g., surveilling populations for punitive purposes).
Documenting key dimensions of metadata (data about data) so that data can be used legally and ethically: • Description of the dataset, its purpose, who created it, etc. • Provenance, or the history of the data, where it came from, why it was collected, and timeline of changes. • Technical specifications that may be needed to use the data, such as file type, format, or software requirements. • Rights related to data ownership, how data may be used, copyright and licenses, and restrictions on sharing and access. • Preservation, or the steps to protect, store, maintain, and back up the data. • Citation information that allows others to properly reference the original source.	Failing to assess, document, and mitigate data integrity issues (e.g., inaccuracies, missing data, values out of range, duplicate rows) that compromise the data's trustworthiness and usefulness for decision-making.
Ensuring that the people whose data are collected understand the purpose, benefits, and risks.	Not providing clear opportunities to opt out of sharing data before, during, and after data collection.
Finding out why people "opt out" of providing data for surveys and other data collection efforts, and using their feedback to minimize harm in future data collection processes.	Failing to consider which data carry elevated risk of harm (e.g., resident HIV status collected by a housing program) and overlooking ways to mitigate risks.
Seeking new data, new measures, and new ways of understanding to drive action toward equity, even when there are significant barriers.	Being complacent, not seeking to improve data collection practices, especially when there are concerns of bias, data quality, and missingness.
Co-creating with community members a framework for the collection and use of RELD/SOGIE data that reflects the overarching mission of the integrated data system and specific community context	Defaulting to national data RELD standards without validating against community needs or avoiding use of RELD/SOGIE data at all because they don't exist or aren't complete.

RELD & SOGIE Data Standards Framework

Collecting primary or secondary data equitably requires engaging with those represented in the data. It is important to understand both how they define themselves and how the data defines them. Collection instruments, both created for the project or borrowed from others, are often imperfect representations, so involving the people behind the numbers provides critical context.

At its best, a framework co-created by community for collecting race, ethnicity, language, and disability (RELD) data and sexual orientation, gender identity, and expression (SOGIE) data both *shows* how systemic and historical racism, oppression, and identity-based power differentials manifest and *compels* change. It tells us where healing needs to happen.

To that end, this guide does not prescribe an answer to the common questions we hear, like "What RELD categories should I use?" or "What are the officially recommended check boxes?" Rather, we offer guidance on assembling a team that can ask and collectively answer questions that arise across the data life cycle about how we collect and use RELD/SOGIE data so that these data can be used to create authentic, purpose-built responses to and relationships with communities.

Below, we review considerations and guiding questions for sites developing a RELD/SOGIE framework for data collection and reuse. We have also highlighted considerations for thoughtful and thorough implementation and change management, the focused effort to implement change with emotional awareness, adult learning, and cooperative principles in mind so the change takes root and sustains itself.

Institutional powers often look to available data and national standards first to define their program's RELD/SOGIE framework, with specific needs of the community and locality being treated as an afterthought. While the national standards can serve as a starting point, we instead recommend working with your governance team to first agree upon the minimum set of identities needed to fully, accurately, and authentically describe identity-based patterns of oppression and asset-based opportunities for the project's purpose. Once you know what you need, you can then evaluate available data and how well the national standards and available data align to your needs.

National standards

Recent updates to the federal Office of Management and Budget (OMB) data standards (Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity) will drive demographic reporting for many programs. Importantly, these standards are more inclusive of ethnic identity than OMB's previous iteration. While race and ethnicity are both social constructs, they are distinct, with race referring to physical characteristics and ethnicity to shared history and culture. Both constructs are important to measure even if you create a more customized framework than the OMB standards. Again, we recommend that any RELD/SOGIE framework, even if based on the national standards, leave room for community engagement and feedback. Finally, your framework should not end with instrument design and data collection. Determining how responses are translated to analytic categories is of utmost importance.

Foundations for Community Involvement: Governance of RELD/SOGIE Frameworks

As with any data effort, a diverse governing body of empowered members is essential to an authentic and accurate RELD/SOGIE framework. This body can be the same as or distinct from the broader integrated data system (IDS) governing body, depending on your context. Look, for example, at the "Setting the Table" activity (Companion Workbook 2.II). When determining whether to build a new body to develop a RELD/SOGIE framework, consider the following: Does the assembled group adequately reflect members of the RELD/SOGIE groups in the data? Does the IDS governing body reflect these groups as well? If not, should it?

When determining if a body is "diverse" or "representative," consider the following: Does the group contain members who can speak from the perspective of:

- Major ethnicities and racial subgroups in the community? Those with visible, and invisible, disabilities? (Consider what data sources you are using, including local rather than federal data sources and community partners when determining subgroups.)
- Those who speak languages other than English at home?
- Those born outside of the U.S. and with differing citizenship statuses?
- Those with a variety of gender expressions, identities, and/or sexual orientations?
- Beyond representation, this group will do its best work if it is surrounded by a culture that embodies the Foundations for Community Involvement. Ensure that key leaders who will need to implement this framework are, at a minimum, consulted throughout the process.
- Before beginning, review these sections and consider who is missing and how well the group is prepared for the following:
- Ensuring members are empowered and prepared to speak on behalf of whatever identities or organizations they choose to represent.
- Conducting meetings, proceedings, and other materials in an accessible manner (including transportation, transcription, and translation services as needed).
- Establishing a clear process to co-define, establish, and adhere to decision-making standards.
- If applicable to your context, developing relationships with local Tribal leadership, including review of Working with Tribal and Indigenous Data and decisions as to whether and how leadership will engage with this process.

Planning: Questions to Create Purpose-Based Design

The RELD/SOGIE data framework should reinforce the mission and purpose established in the Planning phase. Through tailored planning and design decisions, data collection and use can illuminate assets, pathways, and opportunities to grow and share power. For instance, a mission that prioritizes statewide analysis to support reporting on disparities to the federal government may choose to emphasize larger categories of individuals with definitions that can conform to federal standards. An IDS with a mission to hand data to a local community with a distinct demographic population may choose to focus on highly disaggregated data categories to ensure that community members are adequately represented.

Ask:

- How do identities—self-identified, self-perceived, and/or externally perceived—interact with the outcomes you're studying? How have legacies of oppression and opportunity influenced these outcomes?
- Are certain identities disproportionately harmed, or are some particularly visible (which could mean over-surveillance) or invisible (erased), in patterns?
- How do multiple identities in a category, and intersections across categories, magnify or mitigate disparate effects?
- What kind of identity source (self-identification or observed, aka "street race"[14]) is important here?
- How can we validate and test answers to these questions with community leaders and members?
- What minimum population thresholds are needed to fully realize the framework's purpose? Frameworks with a focus on more complete disaggregation may have very small—even as low as 1 or 0—minimum standards for disaggregated group size. Frameworks focusing on broader population trends may need higher minimum sizes for aligned results.

Data Collection: Data landscape assessment for data reuse

Once the governance team has developed an initial sense of identity offerings for a RELD/SOGIE framework, we encourage reviews of available data, including alignment (or lack thereof) among datasets. The goal of this stage is to understand whether current data assets meet the purpose described in the previous stage, to identify whether and what kind of supplemental collection or detail could adapt assets to the purpose, or whether the project needs newly collected data.

Ask:

- What categories of identity are available across datasets? How well do they align with the minimum set described in the previous stage?
- How well do the categories align among themselves? How do the different datasets handle multiple responses, write-ins, "other," unknown, blank, and "decline to respond"? How do they handle changing responses over time?
- How do the data differ by report type: Are they self-reported, completed through proxy, observed, and/or imputed? Is this consistent over time and across sets? Do the individual sets change by source type over time?
- What is the completeness, quality, and consistency of the data? How do responses for the same person compare across sets?

---

[14] López, N. (2024). What is street race? Institute for the Study of Race and Social Justice.

- How were the data collected? What context for data collection purpose, use, and guidance among selections was available to respondents? How does that differ among datasets and how might that affect data quality?
- When identities conflict for a person, what data governance rules are in place to select one or multiple categories?

New Data Collection

If existing data are insufficient for the project, the governance team may decide to collect new data, or dedicate time to partner conversations about data limitations and possible avenues for advocacy. New collection instruments should be designed to match the purpose-based design, avoid being duplicative, and place minimal excess burden on respondents.

Ask:

- What resources are available to the team to run this collection? Consider funding, stakeholder time, community capacity, and institutional trust to sustain a substantive engagement process, technology flexibility, and resources, etc.
- What additional voices might be important to describe, determine, and validate minimum options?
- What were the barriers in the current data assets and how will this new collection overcome them?
- How will the teams ensure that voices from underrepresented communities are heard during the collection process and in the resulting data? For example, Native peoples by Tribal affiliation and even as a whole are often grouped with other identities for lack of "statistical significance."

Implementing new RELD/SOGIE data standards

Implementation of new standards will take time. It is important to resource this part of the work to ensure that the new framework is integrated into existing practice carefully and responsively. This effort involves dedicated training and intentional change management.

We recommend considering the following during implementation:

- Assess existing buy-in among leadership and staff who will need to implement the new framework. Some may still oppose the changes or feel left out of the process and may benefit from additional listening sessions to understand and address their resistance and ensure a smooth implementation.
- Inventory systems that the updated framework will affect, paying special attention to those who will need to actively adapt their systems to make the framework changes possible (in other words, who will be changed—and who needs to change?).

- Work with key influencers and leaders to prepare data documentation, training materials, and "rack cards"—simple community-facing postcards with key "what's in it for me" phrases for those with resident-facing roles.
- Revisit and reassess use of the framework, both immediately after implementation training and at regular intervals.

Data Access, Analysis, and Dissemination: Sustainability and adaptability of standards

The governance group will need to reconvene regularly to ensure the sustainability and adaptability of the framework. The group will address evolving needs, such as developing rules for aggregation; guidance on display labels and footnotes or other usage guardrails; handling of "Other"/Blank/Chose Not to Disclose fields; processes for managing conflicting data; and minimum quality and completion standards for reliable data use. The group may also update initial governance and data collection decisions to adapt to emerging use cases or quality concerns.

Ask:

- Review the questions from governance formation: Does the current group—which may include all or some of the team who formed the original framework—adequately represent identities in the data? Is power being shared equitably?
- What data quality, source, and descriptive contextual information must be shared when data are used?
- What use guidelines are recommended or required to minimize misuse, biased conclusions, or data traumatization?
- How will the group monitor the ongoing use of the data, implementation of its guidelines, and opportunities for improvement/adaptation?

Whether new data are being collected or data reuse is being considered, take the time to align the data to the mission of the project. As with all other aspects of the data life cycle, the community you wish to serve should be centered.

WORK IN ACTION:
Data Collection

Albuquerque Area Southwest Tribal Epidemiology Center (AASTEC), Oversample Initiative

Kevin English

The Albuquerque Area Southwest Tribal Epidemiology Center (AASTEC) aims to produce data that are both robust and representative of American Indian and Alaska Native (AI/AN) populations. To that end, AASTEC, in partnership with state and academic entities, conducts an AI/AN oversample project as a component of the New Mexico Youth Risk and Resiliency Survey (NMYRRS). The NMYRRS, like many data collection tools, randomly samples schools and classrooms statewide. In large, state-level datasets, the AI/AN category is often suppressed or combined with other racial/ethnic categories to increase sample size, and data are rarely available at the county or school level. Tribal communities and programs need both granular and robust data specific to AI/AN youth to inform actionable work around youth health and well-being. That's where the oversample project comes in; the project supplements the NMYRRS random sample with survey data from additional schools with high AI/AN student enrollment to ensure that they are more robustly represented in the results. This is just one way the AASTEC supports equity in data collection. In addition to allocating resources to support the administration of youth surveys in Tribal-serving schools, they also conduct analysis, develop reports, and disseminate data and findings directly back to Tribal schools and community partners. Learn more about their work here.

Allegheny County Data Warehouse, SOGIE Data Collection Standards

Ellen Kitzerow

For more than 15 years, the Allegheny County, Pennsylvania, Department of Human Services (DHS) has been implementing policy and practice improvements to support LGBTQ+ youth and families involved in the child welfare system with the goal of ensuring that all children and families have the best outcomes, regardless of race, ethnicity, age, legal sex, sexual orientation, gender identity, or expression. In 2009, community members first asked DHS to evaluate the extent to which services were culturally responsive and affirming to the LGBTQ+ community, at a time when gender and sexuality were not openly discussed in regard to departmental practices. In response, DHS began a collaborative pilot project with the Center for the Study of Social Policy to develop and implement guidelines for collecting data related to sexual orientation, gender identity, and expression (SOGIE). The department knew that data would be essential to monitoring their progress as they sought to train staff and partners to provide more affirming care. But before collecting new information, the project team needed to work closely with providers, system-involved youth, and community members to address privacy and data concerns, and to understand the possible implications of sharing data with external organizations. They also committed to updating their IT system to include more robust privacy protection, mitigate harm, and train staff, including IT and case workers, to ensure careful implementation of new data practices. After the guidelines were field tested, they were broadly adopted and are now standard in family service provision, demonstrating the importance of SOGIE data for improving services across child welfare.

Today, the Office of Children Youth and Families provides SOGIE information (or its absence) at both the aggregate and individual level to help leadership and staff measure efforts and understand continuing gaps. At the same time, DHS continues to monitor providers' willingness to foster LGBTQ+ youth, and shares that information in settings where agencies are addressing issues of retention and recruitment of foster homes. From the beginning of this work, Allegheny County has been committed to protecting the privacy and confidentiality of clients' SOGIE information. To this end, SOGIE information is shared and used only when it is deemed necessary to make decisions around placements and other casework practice. When marked "Confidential" in the case management system, a client's sexual orientation, gender identity, and pronouns are hidden, and only a small number of staff connected to that case have access to that information. The aim is always "to do our best to honor our clients' decisions—in both what they are comfortable sharing and what they are comfortable with us documenting and discussing."

Connecticut, Race, Ethnicity, Language, Disability Data Collection Standards

Sumit Sajnani

Since 2021, the Connecticut Office of Health Strategy (OHS), a core member of the state's Equity in Practice Learning Community team (see State of Connecticut), has led a collaborative effort with a broad range of partners to design and implement state-wide standards for collecting race, ethnicity, language, and disability (REL-D) data. Updated standards are published annually in order to capture changing needs over time. Having these data in a standardized, granular format enables state agencies, provider organizations, and policy makers to identify disparities in service delivery and make significant strides toward a more equitable health system. These standards ensure that data are self-reported with patient consent and that receiving services is not contingent on providing consent. Read more about Connecticut's public act and statute on collecting REL-D data and review their implementation toolkit and other important process documents here.

Mapping Racist Covenants (MRC), Digitization and Indexing of Paper Records

Jason Jurjevich

The Mapping Racist Covenants (MRC) project maps the covenants, conditions, and restrictions (CCRs) put in place in Tucson homeowners' associations and subdivisions from 1912 to 1968, alongside more recent census data from 1960 to 2020. The map visualizes the impact of racist CCRs on contemporary segregation and housing inequity. To create the map, researchers had to first identify properties developed during the years when racist covenants were proliferating, then visit the local Records Office to access and digitize the original records for each area. The team mapped their findings to show which subdivisions contained racist covenants as well as logged the groups mentioned in each covenant and identified whether it was amended, canceled, or enacted "in perpetuity." Now, anyone can explore the map on a user-friendly website to see if a subdivision has or once had a racist covenant, and can access an archive of the CCR's original language for each affected subdivision.

Though racist CCRs have been illegal for over 50 years, MRC asserts that they remain a critical, and largely unexplored, barrier to housing equity. Of the 121 subdivisions where racist covenants are "[enforced] in perpetuity," many remain more than three quarters White to this day. By digitizing, indexing, and mapping the extent of their impact, MRC enables Tucson residents to better understand the roots of inequities in their communities, and enables policy makers to explore opportunities to ameliorate harms. This work also models the importance of archival work in understanding and responding to modern challenges. View the map here and learn more about the movement here.

Oregon Health Authority (OHA), Oregon Behavioral Health Survey

When the Oregon Health Authority (OHA) noticed underrepresentation and inaccuracies in the behavioral health data of communities of color, they embarked on a process to redesign the way this information was collected. OHA staff partnered with community members closest to the issues being researched to evaluate shortcomings and co-construct new questions to be asked as a part of the Behavioral Risk Factor Surveillance System. Culturally specific workgroups were also formed and facilitated by a local coalition group trained in research justice frameworks. Through this structured and community-guided process, OHA was able to incorporate community voice and perspective to ensure that data collection was more representative. The partnership also enabled increased community engagement in the collection process. Learn more about their process and results here.

Resources: Data Collection

Data Disaggregation Action Network Resources, The Leadership Conference Education Fund

Interagency Committee on Disability Research, 2024, Administration for Community Living

Learning from Indigenous Ways of Knowing, 2021, We All Count

Measuring Sex, Gender, Identity, and Sexual Orientation, 2022, National Academies of Sciences, Engineering, and Medicine

More than Numbers, 2020, Charles and Lynn Schusterman Family Philanthropies

Our Identities, Ourselves: A Guide to Anti-Racist Data Collection for System Leaders and Data Administrators, 2021, The Center for the Study of Social Policy

Racial Bias Assessment Tool, 2021, Chapin Hall

Racial Equity in Data Access

Data access generally refers to practices regarding who can securely obtain, view, or use data; when it can be used; and for what purpose. We recommend that agencies and community-based organizations use a multi-tiered approach to data access that begins with a clear delineation of practical and legal data availability.

OPEN DATA	PROTECTED DATA	UNAVAILABLE DATA
Data that can be shared openly, either at the aggregate or individual level, based on state and federal law. These data often exist in open data portals.	Data that can be shared, but only under specific circumstances with appropriate safeguards in place.	Data that cannot or should not be shared, either because of state or federal law, lack of digital format (paper copies only), or data quality or other concerns.

The determination about which data are open, protected, or unavailable can have significant implications for equity. The intended and unintended consequences of a data release or the withholding of data may disproportionately impact certain individuals or communities. When making decisions about data access, it is essential to have strong governance in place. Governance will help determine whether making the data open is valuable and mission aligned or, in the case of protected data, whether those who request access are using the information in ways that are mission aligned and beneficial to those represented in the data. Good governance protects those represented in the data and has clear structures to prevent data access from relying on transactional relationships.

Open data are most often available online, through search queries, static PDFs, CSV files, and front-end data visualization tools (such as dashboards). Open data is not inherently good or bad, equitable or inequitable. Making data more open and democratizing access to information can support equity efforts (see MAREA, Strong Start Index, and Black Data Wealth Center). Making data more open without thoughtful interrogation, context, and presentation can also mislead users, contribute to pre-existing biases, and harm individuals and communities. Providing the appropriate contextual metadata, as well as information around data release schedules and the process for requesting that data be released, is an important step (see Child Opportunity Index and Justice Outcomes Explorer). We encourage collaborative governance to assess and balance the risks and benefits of whether and how to publish specific datasets and create an equitable and ethical open data environment in your context.

Protected data are generally identifiable information collected for operational purposes (e.g., for caseworkers to manage service provision); however, these data can be useful for researchers, community leaders, and other constituencies outside of direct service agencies. There is a delicate balance between hoarding high-value data that could be used to improve lives, and risking individual or community harm if data are shared or used inappropriately. While working to strike this balance, consider how data are processed, managed, and requested (see NCDHHS, Charlotte Regional Data Trust, and KYSTATS). Ideally, when protected data can be beneficial to those in the data, there should be pathways for it to be accessed (see Client View and Built for Zero). Once again, data governance is how we build those pathways with care.

Unavailable data are data that cannot or should not be shared because access by external users is prohibited by law (e.g., HIV status); information is not yet digitized (e.g., eviction records); there are substantial data quality issues (e.g., data are missing or inaccurate); or the data have been deemed too sensitive for any release in case of redisclosure (e.g., information on domestic violence). There are many legitimate reasons that data are unavailable, and those reasons should be documented to ensure that valuable data are not being hidden or made unavailable for illegitimate reasons (e.g., technical problems).

Categorizing data is a large task, but it is critical to centering racial equity. The risks and benefits of making data open, protected, or unavailable should be carefully considered. The risk vs. benefit matrix and the positive and problematic practices detailed below can be helpful tools in making and communicating about these categorization decisions.

Note: Considering privacy and security is necessary for data classified at all levels. The practices below, as well as the Work in Action, highlight some strategies to maintain privacy without gatekeeping valuable information (see BEST and Urban Institute).

Positive and Problematic Practices: Data Access

POSITIVE PRACTICE	PROBLEMATIC PRACTICE
Privacy & Security for All Data Types	
Having a comprehensive privacy strategy as the foundation for decisions, policies, training, and procurement processes.	Approaching privacy in an ad hoc fashion.
Balancing data access and data privacy by adhering to best practices for managing and protecting sensitive data (e.g., "Five Safes" framework).	Assuming that best practices for data management and security are being followed without explicit protocols, oversight, and training in place.
Carefully discerning how de-identification and anonymization of data will take place prior to release.	Releasing data that can be re-identified by individuals (e.g., publishing small geographies or cell sizes) or by more advanced data mining techniques.
Evaluating privacy, security, data quality, and other risks relative to the benefit of providing data access, and engaging community members in the data governance process.	Releasing data without proper consultation with data owners (e.g., Tribal Consultation is required whenever data that could identify a Tribe or Tribal members are disclosed).
Open Data	
Deciding which data to make publicly available based on value and benefit to community partners and those represented within the data.	Releasing open data based on indices, algorithms, or other sources with a history of discriminatory impact on communities (e.g., "teacher effectiveness scores" and "school report cards").
Disseminating information that enables community members to effectively use open data (e.g., contextual information, data release schedules, how to access data, process for requesting new open datasets and how requests are evaluated).	Assuming that the data presents a full picture of reality and not publishing contextual metadata (e.g., source; why, how, and when it was collected; its relationship to structural factors; data quality considerations; whom to contact for questions and clarification).
Protected Data	
Outlining clear data request policies and procedures for all requesters, both internal and external.	Failing to refine data request processes based on user experience and solicited feedback.

Sharing information that helps requesters submit a high-quality data request (e.g., metadata for available datasets and variables, research/policy priorities for data use, evaluation rubrics for data requests).	Allowing data requesters to circumvent established processes for accessing data, whether intentionally or unintentionally.
Establishing a clear pricing structure for data access and using it consistently (e.g., flat fee, hourly rate for staff effort, discounts for student- or community-led projects).	Using a one-size-fits-all pricing model without the input of governing bodies.
Finding legal pathways to share protected data when it can be used to improve programs, services, or people’s lives.	Hoarding high-value data under the guise of legal restrictions or by limiting access only to those with insider connections.
Sharing data to reduce administrative burden on clients and communities (e.g., referral coordination, streamlined eligibility determination).	Providing more data than is useful or necessary, instead of curating and sharing variables based on what is needed and of sufficient quality for the project at hand.
Utilizing privacy-preserving technologies as a tool to develop cross-agency relationships and build trust.	Allowing privacy-preserving technologies to replace participatory governance.
Incorporating conflict resolution approaches within data governance policies to handle disagreements regarding data access.	Refusing to share information with crucial constituencies and community leaders (e.g., Tribal Nations, community organizers).
Unavailable Data	
Clearly documenting why data are unavailable (e.g., specific statute, legislation, data quality concerns, data not digitized, undue burden in data preparation).	Storing potentially valuable data without creating pathways for data access (i.e., data mausoleum practices).
Utilizing privacy-preserving technologies to mitigate risk in sharing and integrating protected data, or even to allow for high-impact use cases of sensitive data that would otherwise be classified as unavailable.	Failing to consider and implement advancements in technical tools that can enhance privacy and security.
Protecting the rights of data owners to access their data by building it into data management practices and contracts.	Having data that are unavailable for technical reasons rather than a legal or other legitimate restriction on access.

WORK IN ACTION:
Data Collection

Open Data

Neighborhood Nexus, Metro Atlanta Racial Equity Atlas (MAREA)

Nikolai Elneser & Tommy Pearce

The Metro Atlanta Racial Equity Atlas (MAREA) is a digital platform that showcases how local systems interact and affect neighborhood and individual experiences differently based on race and ethnicity. It combines storytelling, interactive maps, and calls to action while equipping users with advocacy and decision-making tools. MAREA was designed to be (1) impactful beyond the initial release of the tool, (2) focused on assets more than deficits, and (3) actionable, not just shocking.

MAREA was developed to allow users to interact and explore different barriers through an open dataset and visual tools that emphasize how systems and historic inequities produce disparate community outcomes. To show how race interacts with the different systems, an overlaid black outline on top of the socioeconomic indicators represents areas with a majority of residents of color. The overlay shows that barriers to opportunity are racialized, without creating a deficit narrative. Additionally, the accountability tools section offers a central repository for decision-making resources connected to ongoing advocacy efforts. The platform offers in-depth analysis to help decision-makers and communities use the data to focus their efforts more strategically toward equity.

Diversitydatakids.org, Child Opportunity Index (COI)

Dolores Acevedo-Garcia & the diversitydatakids.org team

The Child Opportunity Index (COI) is a composite index of neighborhood features that help children thrive, capturing variation in opportunity across U.S. neighborhoods and over time. The COI is an efficient means of leveraging multiple information points from data that already exist, framed to center racial equity. The index does not include race as an indicator, choosing to use indicators for which race is often a proxy. The COI allows for strategic targeting of resources to where they are most needed, and the team works hard to maintain the system and support users. The COI is both a data tool and a data intermediary because the team makes themselves accessible and available to help solve problems or answer questions for those using the index. Their accessibility and clear documentation make it easy for many different users to engage and use the COI in making strategic decisions and better understanding geospatial impact. Explore the index here.

Criminal Justice Administrative Records System, Justice Outcomes Explorer (JOE)

Jordan Papp & Chandler Rombes

The Justice Outcomes Explorer (JOE) is a publicly available data dashboard that provides a look into how the criminal justice system touches the lives of millions of Americans every year. Using data from the Criminal Justice Administrative Records System at the University of Michigan linked with other records available from the U.S. Census Bureau, this tool offers novel statistics on justice-involved individuals' outcomes that more fully represent their life trajectories, including characteristics related to recidivism, employment, earnings, health, and public assistance. The landing page displays example questions and points users to the appropriate queries to answer them. Next, the data stories page features case studies to give users ideas on how to fully leverage what's available. The explorer also allows users to create different visualizations, showing data in context, data over time, or data maps. Finally, the "get the data" page allows users to download the full set of statistics that power JOE. The tool's user-friendly design makes the data and meaning-making process more accessible. Explore the data here.

University of Southern California, Children's Data Network, Strong Start Index

Regan Foust

The Strong Start Index is a population-based tool that uses information on birth records to characterize, in a standardized way, the conditions into which babies are born within and across California communities. With an overall goal of supporting health, equity, and resilience, the index takes a strengths-based approach, combining 12 indicators present on each child's anonymized birth record that are standard across years, populated consistently, and theoretically and empirically related to good outcomes for children throughout the lifecourse. When scores are summed at the child level, aggregated to the census tract level, and then visualized in various ways and for populations of interest, information on birth records can further our understanding of the distribution of resources among California communities, highlight disparities, and facilitate more equitable investment. In the coming years, the Strong Start Index will be used to develop a population-based sample of parents who recently gave birth in California. This sample of families will be asked directly about service needs and experiences, and for consent to link their service records going forward. This next step promises to fill a major gap in the literature about newborns and their families and make our efforts to improve children's outcomes more intentional and equitable. Explore the Index and learn more here.

Black Wealth Data Center (BWDC)

David Asiamah

The Black Wealth Data Center (BWDC) leverages the power of data to view the Black experience through an equity lens. Launched and supported by Bloomberg Philanthropies' Greenwood Initiative, BWDC provides one of the most comprehensive interactive databases for racial wealth equity, helping public and private sector leaders access actionable data for policy making and investment decisions to address disparities. BWDC aggregates extensive, high-quality datasets from federal agencies such as the Census Bureau and the Federal Reserve and adds additional depth by bringing in complementary data from private and local sources—all available in one place for users to access quickly, efficiently, and at no cost. Their website provides data visualizations designed with clarity and interactivity in mind. Users can explore various indicators or download the data to use for their own reporting and analysis. A comprehensive methodology page notes when and why data were collected and where each metric can be accessed on the site. With a focus on reliability and transparency, BWDC aims to build trust as a go-to data resource for organizations and researchers seeking a deeper understanding of wealth inequities in the U.S. to advance innovative solutions promoting Black wealth.

More Open Data Examples:
Baltimore City Youth Opportunity Landscape, 2023, Baltimore's Promise
Chicago Equity Dashboard, Chicago Health Atlas
The Opportunity Atlas, U.S. Census Bureau

Protected Data

Allegheny County Data Warehouse, Client View and AccessMyInfo

Kathryn Collins

The Allegheny County Data Warehouse, hosted by the county's Department of Human Services (DHS), brings together and integrates client and service data from a wide variety of DHS-administered programs, including child welfare and behavioral health, as well as other government agency data like public benefits and court involvement. The Data Warehouse design allows the use of integrated data to improve service delivery through Client View—an application for case managers to assess a client's claims and service utilization across DHS information systems, programs and time. Privacy, confidentiality, and ethical use are core components of training for workers who utilize Client View. Credentialed logins and search approvals are among the multiple safeguards in place. More information on the privacy/confidentiality training and a full list of the information available through Client View is available here.

Service recipients can also see where and how they are represented in the data through AccessMyInfo. The platform allows residents to understand what information is collected about them, and provides them a view into their own service history within Allegheny County. Detailed information about what records are on file and how long records are kept is available here.
The Allegheny County Data Warehouse is among the oldest and most advanced data integration efforts in the U.S. It is one essential arm of an agency committed to ongoing community engagement and transparent decision-making across their programs and operations. Learn more about this work in AISP's 2014 case study.

Birth through Eight Strategy of Tulsa (BEST), Spotlight Pilot

Erin Powell & Aaron Bean

Spotlight is a socio-technical approach to rapid, actionable, and secure data sharing, developed by Asemio through a pilot project with Birth through Eight Strategy of Tulsa (BEST) and seven local social service agencies. Spotlight uses secure hash encoding and privacy-preserving record linkage to ensure that personally identifiable information remains at the source—never moving from one organization to another—while still enabling individual-level linkages across datasets. This is not only more efficient than traditional data sharing, but allows Spotlight users to disaggregate data to reveal and interrogate patterns of inequities without compromising the privacy of the individuals in the datasets, even in small areas.

The BEST pilot allowed partners to share over 12 years of data from 32 programs in under two months. Since the 2019 pilot, this process has been used to share data and gain insights across a vast array of social priorities such as PreK-12 education, higher education, workforce development, food access, eviction prevention, justice reform, and early childhood health and well-being. It has opened access to state-level data (e.g., Medicaid enrollment patterns) that has been used to address service gaps among local nonprofit providers in ways that were not previously possible. Finally, it has served to bolster trust between community partners and increase engagement in collaboratively sharing, interpreting, and using data to inform service strategies.

Since the initial pilot project, Spotlight has been used for over two dozen subsequent projects. To read more about the BEST Spotlight pilot project, download the case study here.

Built for Zero, By-Name Data

Adam Reuge

Built for Zero is a national movement and methodology to end homelessness through collection and conscious sharing of by-name data. By-name data refers to the collection of identifiable information about individuals experiencing homelessness, with consent. Specifically, by-name data capture who is experiencing homelessness in real time to help answer key questions such as how many people are becoming homeless for the first time; how many people are returning to homelessness; how many people are no longer homeless; and whether experiences of people moving through the system are equitable. Further, data with individual names can be used for case conferences so that community providers can better match services with each individual's needs. Over 140 communities are using the Built for Zero system to legally and ethically challenge the systemic drivers of homelessness and improve lives. Learn more about how this approach helped Rockland, Illinois, become the first community in the country to reach "functional zero" for both veterans and people experiencing chronic homelessness here. Learn about the benefits of by-name data here.

Charlotte Regional Data Trust, The State of Our Data Report and Data License Request Portal

The Charlotte Regional Data Trust is a partnership that links data across service and organizational silos to make connections visible and provide information that the community can act upon. Building upon robust governance structures, the 2024 State of Our Data Report provides communities, partners, and interest holders with an accessible summary of this enduring infrastructure, the data that is held in trust, and why this matters. This report includes an overview of data sources that can be broken out by topic, geospatial and organizational representation, and descriptive statistics of the individuals and communities represented within the data system.

The report and newly developed Data License Request Portal was presented during the 2024 annual Board of Directors meeting, an in-person event with 250+ people in attendance, including the Board, staff, data partners, community members, and the members of five committees and workgroups. This annual meeting included a demonstration of how data integration across Data Trust assets is conducted and presented the newly released Data License Request Portal, which includes codebooks, documents, and a more automated and efficient data license request process.

Kentucky Center for Statistics (KYSTATS), Data Access and Use Policy

Matt Berry

The Kentucky Center for Statistics (KYSTATS) is a state agency that collects and links education and workforce data from 27 agencies and 47 sources for use by policy makers, practitioners, and the public. Their protocols and processes emphasize transparency, ease of use, streamlined access, and privacy protection. Through KYSTATS.ky.gov, users can access various reports, blog posts, or dashboards, or make specific data requests.

Staff work closely with government, school, and business partners to continually assess report effectiveness. The KYSTATS Data Access and Use Policy provides context for how agencies, residents, and providers access data based on their user type. KYSTATS uses a cost-recovery model to remove access barriers between providers and the data that they need. Published flow charts articulate the data request process and response timeline. Additionally, the KYSTATS data use dictionary aims to close the gap between data professionals and the public by providing context around which data are and are not available for use. Finally, the organization's clear anonymization protocols are aligned with industry best practices to avoid redisclosure and protect data privacy. Learn more here.

North Carolina Department of Health & Human Services (NCDHHS), Data Sharing Guidebook

Amy Hawn Nelson

North Carolina Department of Health & Human Services (NCDHHS) regularly shares data across the department, as well as with other local- , state- , and federal-level government agencies and research partners. While both sharing and integrating data are common practices to support NCDHHS' operational goals, historically this work has been time- and labor-intensive. From 2019 to 2024, AISP partnered with NCDHHS to conduct a data landscape and participatory action research to understand how to more effectively share and integrate data. These learnings were captured in a co-created Guidebook, which has supported improved processes and resulted in significant time reductions to facilitate enterprise-level data sharing. The NCDHHS Data Sharing Guidebook, updated annually, establishes clear pathways for data sharing and integration and establishes a common legal framework for data sharing and integration that supports more efficient and secure use of data. Learn more here.

Unavailable Data

The Urban Institute, Creating a Fully Synthetic Data Set

Gabe Morrison, Geoffrey Arnold, Kathryn Collins, Joanne Foerster, David Walker, Claire Bowen, Robert Gradeck, & Alena Stern

Allegheny County Department of Human Services (DHS) holds highly protected data, which is typically restricted, suppressed altogether, or released only in aggregate. While these measures protect individual privacy, they can also limit data use for research. In 2023, DHS developed a *fully* synthetic dataset in order to both preserve protections and dig deeper into relevant questions about service impact, such as interaction effects between identities, overlaps in program participation, and continuation of service use from month to month. Synthetic data replace actual records in a dataset with "pseudo-records" generated from statistical models to be representative of the original records. Unlike a partially synthetic dataset, a fully synthetic dataset has no link to the underlying confidential records, so there is less risk of re-identification. This allows researchers to detect patterns from the confidential data while reducing the risk of privacy violations, making data that would otherwise be unavailable accessible and actionable. Read more about the fully synthetic dataset here.

Resources: Data Access

Data User Guides, 2024, Western Pennsylvania Regional Data Center

Disaggregation of Public Health Data by Race & Ethnicity: A Legal Handbook, 2022, Network for Public Health Law

FAIR Principles, 2016, GOFair

Five Safes Framework, 2017, Office for National Statistics, UK Data Service

North Carolina Department of Health & Human Service Data Sharing Guidebook, 2024, NCDHHS and AISP

Open Data Handbook, 2024, Open Knowledge Foundation

Race and Ethnicity Data Guidebook, 2024, DC Mayor's Office of Racial Equity

Racial Equity in Data Analysis

Data analysis is the stage at which available data are explored in order to develop findings, interpretations, and conclusions. Data analysis can be as simple as calculating descriptive statistics, such as counts of program participants or the percentage of participants who achieved a certain outcome. Analysis can also include measuring longitudinal trends, identifying causal relationships between interventions and outcomes, or creating complex models that predict participant behavior.

Incorporating a racial equity lens during data analysis starts with having the right mix of people to develop and execute a strong analytic plan. This includes subject matter experts with deep understanding of the existing evidence and most relevant questions to ask (see Indiana MPH in the Work in Action); the "data people" who know how to clean and wrangle the available data, assess data quality, and apply proper statistical methods; and, when appropriate (see Foundations for Community Involvement), community members with lived experience of the issue being studied who can ensure that analytic approaches are aligned to community need and support the interpretation of results (see Northside Achievement Zone and Wilder Research).

Data analysis involves a complex series of decisions about the questions being asked, the data and methods used to answer them, and how results will be interpreted to inform conclusions and recommendations. Decisions about **data disaggregation,** in particular, require careful consideration. On the one hand, disaggregating data can shed light on the unique experiences of small populations and those glossed over in other analyses. However, creating a subgroup has implications, and may shift the focus of analysis to a specific population that is already over-surveilled. Another key area of decision-making in analysis is whether to use **quantitative data, qualitative data,** or a **mixed methods approach.** We find that too often, the opportunity to strengthen quantitative analysis by weaving in qualitative data and other forms of contextual data, such as the social and political history of race in the local area (see Hacking Into History and Delaware PDG), is overlooked. Finally, decisions about how you frame and tell the story of the data matter (see CalEnviroScreen 4.0 and Library of Missing Datasets). Though we will cover these decisions more in the section on Racial Equity in Reporting & Dissemination, it is worth noting that **framing** begins in the analysis stage, and solely relying on statistical outputs will not necessarily lead to insights or empower people to take action. Across the board, engaging with individuals who have lived experience and with trusted community advocates can strengthen these decisions, leading to more meaningful and robust analyses.

Positive and Problematic Practices: Racial Equity in Data Analysis

POSITIVE PRACTICE	PROBLEMATIC PRACTICE
Engaging a range of expertise (e.g., subject matter, methods, lived experience) to ensure that the data analysis approach is appropriate for the research questions and local context.	Not building in time and resources for data discovery and exploratory data analysis before diving into formal analyses.
Making sure the people responsible for data wrangling understand the datasets, variables, and analytic plan.	Failing to document potential data quality issues, processes used to remedy issues, and how these may impact analyses.
Designing mixed methods analytic plans that purposefully seek out and combine data sources to better understand social problems through "multiple ways of knowing."	Creating analytic approaches that are indecipherable to nonexperts without explaining them clearly for a general audience.
Disaggregating data and analyzing intersectional experiences without compromising data privacy.	Assuming that data representing small populations are not meaningful to analyze because of statistical insignificance.
Training analytic staff on best practices for analyzing RELD and SOGIE data (see RELD & SOGIE Data Standards Framework)	Failing to recognize the distinctiveness of identities and intersectional experiences (e.g., assuming gender queer youth, youth of color, and gender queer youth of color all have similar reasons for program nonparticipation).
Carefully considering how subgroups are defined, analyzed, and reported, with an emphasis on asset-based framing.	Not using appropriate comparison groups to contextualize findings (e.g., assuming White outcomes are normative).
Highlighting structural factors within analyses (e.g., overlaying redlining data to correlate place to outcomes).	Using one-dimensional data to propel an agenda (e.g., use of student test scores in isolation from contextual factors such as teacher turnover or school-level demographics).
Drawing on community member expertise when interpreting analyses and identifying root causes of findings.	Interpreting results without examining larger systems, policies, and social conditions that contribute to disparities in experiences and outcomes (e.g., poverty, housing segregation, access to education).
Empowering professionals and community members to use the results of analyses to improve their work and their communities.	Analyzing data with no intent to drive action or change that benefits those represented in the data.

WORK IN ACTION:
Data Analysis

Indiana Management and Performance Hub (MPH), Coordinated COVID-19 Response

Josh Martin

The Indiana Management and Performance Hub (MPH) played an essential role in early COVID-19 prevalence and spread research, as part of a project led by the Indiana Department of Health and researchers at the Fairbanks School of Public Health at Indiana University–Purdue University Indianapolis. When the COVID-19 pandemic first reached the U.S., the understanding of infection rates was limited to positive test rates when what was needed was a full picture of the proportion of the population these positive tests accounted for. MPH was able to leverage their infrastructure and trusted relationships with government agencies to create a population-level dataset that allowed researchers to understand the prevalence of COVID outside of symptomatic cases and hospitalizations. Using tax records, MPH and Fairbanks were able to randomly sample the population and build what became the "denominator" for COVID-19 analyses across the state and nationally. Indiana went on to test over 15,000 residents in April 2020, and MPH continued to serve as an integral partner, supporting vaccine distribution and accelerating public health response to mitigate disparate impacts of the pandemic. For example, after researchers found there was a 90 percent vaccine appointment vacancy rate in a high-risk neighborhood, MPH was tasked with providing a contact list of people within that zip code, and, within one hour of targeted communication, vaccination appointments there were 100 percent full. Read more in the full AISP case study here.

California Environmental Protection Agency, CalEnviroScreen 4.0

The CalEnviroScreen 4.0 Mapping Tool (CES) identifies California communities most impacted by multiple forms of pollution and visualizes where disproportionality and overburden are experienced. A story map used in conjunction with CES explains the relationship between race/ethnicity and CES results, beginning with how the practice of redlining in the 1930s not only forced people of color to live in more polluted areas but also increased the amount of pollution in those same neighborhoods. When researchers compared Home Owners Loan Corporation designations and CES vulnerability designations, they found that on average redlined areas had higher vulnerability scores. This important context highlights the role of systemic factors in the contemporary disadvantage of communities of color. By showing the role of government action and inaction, CES helps to remove blame from contemporary communities burdened by the effects of a century-old policy. Learn more about CES 4.0 here and see the full story map here.

UNC Chapel Hill & NC Data Works, Hacking into History

Alex Chassanoff

Hacking into History is a community-driven project that explores the story and impact of racist covenants found in property deeds in Durham, North Carolina. The project began in March 2020 and is a unique collaboration between legal scholars, county government, civic data intermediaries, experimental artists, community facilitators, and archival studies scholars. Through regular workshops and working sessions on Zooniverse, a citizen science platform, volunteers were trained to identify, transcribe, and validate property deeds containing racial covenants. The project analyzed over 200,000 property deeds written between the 1880s and 1973, and trained over 300 volunteers to identify, transcribe, and validate property deeds containing racial covenants. These primary source historical records had been previously digitized but were yet not in a usable format for analysis. Machine learning was used to initially identify racist covenants, with human review. Most importantly, the volunteers were supported through skilled facilitation while engaging with explicitly racist language. This portion of work concluded in October 2023, with ~2,800 deeds verified to contain racial covenants. The team also created the Hacking into History Curriculum Kit, which contains pilot exercises, sample deeds, a slide deck, and additional resources that K-12 educators can use for teaching students about racial covenant clauses in Durham. Since that time, the project has shifted into a community of practice model. A small group of neighbors and citizens meets monthly to hold conversations about the ongoing impact of these covenants (and other exclusionary housing practices) and to consider next steps for action. Read more about the project in case studies from Civic Switchboard and Society of American Archivists.

Northside Achievement Zone (NAZ), Understanding NAZ Families' School Choices Through Data and Storytelling

Amy Susman-Stillman, Briellen Griffin, Piere Washington, & Jessie Austin O'Neill

In January 2024, the Northside Achievement Zone (NAZ) and Wilder Research, NAZ's evaluation partner, piloted a meaning-making session as part of a larger effort to strengthen an equitable data culture at NAZ and increase families' use of data to gain insights into themselves and their journey toward success. NAZ hosted families, including parents, caregivers, and their scholars, to enjoy food, fellowship, and discussion about data collected during the 2022-2023 school year related to schooling choices. Since 2010, when NAZ began, enrollment of NAZ families in anchor schools (i.e., home attendance boundary) has declined, meaning that fewer scholars are attending NAZ partner schools, which suggested instability in the local school choice framework. Together, the group made meaning of the data and insights that emerged from the 2022 NAZ Family Needs Survey about navigating school choice.

In the 75-minute session, parents and caregivers shared their stories alongside staff, weaving together a complex narrative of diverse school choice experiences. Through the context of storytelling, families felt comfortable disagreeing, allowing the data to reflect a variety of experiences. The session helped dispel the myth that Black lived experiences are monolithic. The families expressed the need for more peer support, particularly by ages and stages (e.g., early childhood, elementary, secondary, and post-secondary). The NAZ team now has a better understanding of the nuances of individual experiences and came away with new support strategies to consider.

Delaware Preschool Development Grant, User Experience Needs Assessment

Caitlin Gleason

Recognizing the barriers to early childhood care and education access in their state, Delaware, EY-Parthenon, and IDEO conducted the Delaware PDG B-5 User Experience Needs Assessment to understand and center the experiences of parents with young children. The research design included family and professional interviews, family observations, and pop-up prototyping events at playgrounds and state service centers. By going to where families live, work, and play, researchers could understand not just what families say, but also the "natural inconsistencies of everyday life." The report findings highlight a core issue with the childcare system, that parents want to feel cared for and welcomed—a conclusion that could not have been reached without expanding the analysis beyond quantitative sources to incorporate family experiences with systems, both historical and current. Read the whole report here.

The Library of Missing Data Sets

Mimi ỌNỤỌHA

The Library of Missing Datasets is an ongoing physical repository of things that have been excluded in a society where so much information is collected. Empty folders are each titled with the name of a missing dataset, all taken from a comprehensive list that the artist has been collecting since 2015. Examples include "people excluded from public housing because of criminal records" and "How much Spotify pays each of its artists per play of song." The work calls attention to a common and implicit message: that if there is nothing to see, it is considered unimportant. In the words of the artist: "That which we ignore reveals more than what we give our attention to. It's in these things that we find cultural and colloquial hints of what is deemed important. Spots that we've left blank reveal our hidden social biases and indifferences." Making missingness tangible in this way gives uncollected information credence, and offers a way to understand the power of absence.

Boston University Center for Antiracist Research, COVID Racial Data Tracker and COVID Stories

Ibram X. Kendi & Dawna Johnson

The COVID Racial Data Tracker—a collaboration between the COVID Tracking Project and the Boston University Center for Antiracist Research—highlights how people of color have been disproportionately impacted by the COVID-19 pandemic. While active, the tracker provided the most complete race and ethnicity data on pandemic impacts and was used to advocate for more county-level reporting of these categories. In an effort to add further context to the data, the Center for Antiracist Research collected videos, photos, and essays from people of color to share how COVID-19 affected their lives and the people around them. Individuals from across the U.S. shared their stories, including topics like navigating job loss, managing the mental health effects of lockdown, and engaging grief practices during times of widespread loss. The stories show how data points are just one way of understanding how communities responded to and were changed by the pandemic.

Minnesota, Missing and Murdered Indigenous Relatives (MMIR) Task Force Report

Nicole MartinRogers

The Minnesota Missing and Murdered Indigenous Relatives (MMIR) Task Force was created to examine the root causes of, systemic problems of, and potential solutions to violence against Indigenous women and girls, including members of the Two-Spirit community, who make up a disproportionate number of victims. Established by the Minnesota Legislature and signed into law by Governor Tim Walz in 2019, the MMIR Task Force included representatives from all 11 Tribal Nations in Minnesota, community and advocacy organizations, legislators, law enforcement, and the legal field. Wilder Research assisted the Task Force by conducting extensive background research, facilitating public hearings and comment sessions across Minnesota, and supporting the development of the data-informed mandates. The Task Force's final report recognizes that jurisdictional complexities require that solutions be tailored to the needs of the community, especially centering the sovereignty of Indigenous territory. The Report's first mandate, which was adopted by the

Legislature in 2021, was to create an MMIR Office in recognition that this is an ongoing problem. The MMIR Office continues to work with the local community to use both quantitative and qualitative data to drive action toward justice.

Resources: Data Analysis

Recommendations from the Equitable Data Working Group, 2021, Equitable Data Working Group

Five Ethical Risks to Consider Before Filling Missing Race and Ethnicity Data, 2021, The Urban Institute

Community Engagement in Causal Analysis Part 1 & Part 2, 2024, We All Count

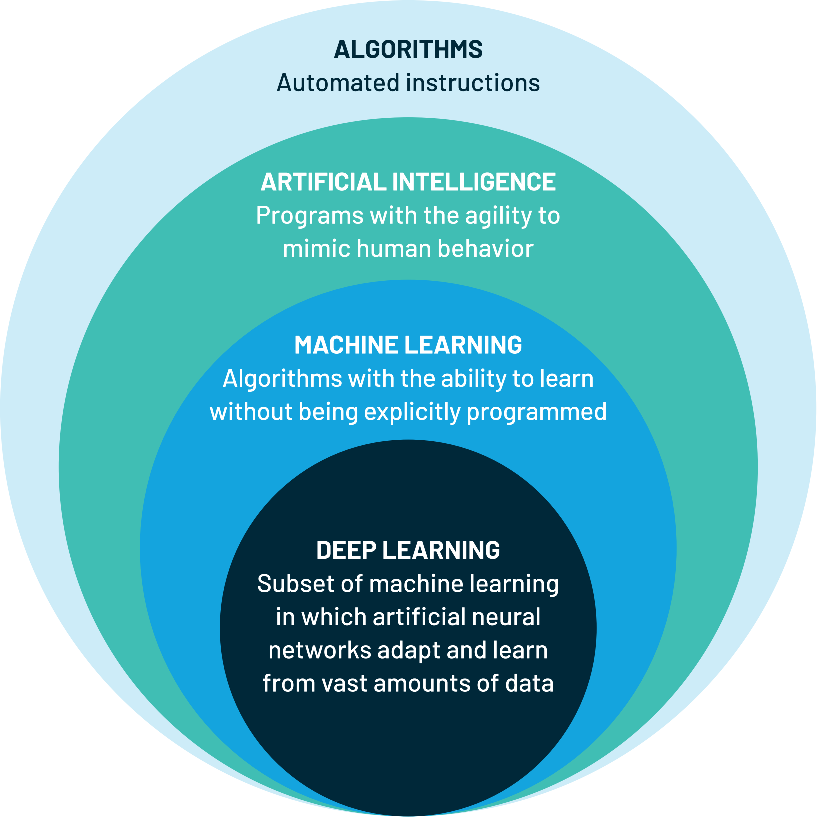
The SEEDS of Indigenous Population Health Data Linkage, 2021, Rowe, Russo Carroll, Healy, Rodriguez-Lonebear, & Walker

Racial Equity in Use of Algorithms & Artificial Intelligence

The proliferation of publicly available algorithmic tools, including artificial intelligence, is rapidly changing our relationship with data, generating new ethical questions and considerations for the public sector. **Algorithms** are statistical tools that allow for automated problem solving. These tools can put together data in new ways that have the potential to be leveraged for social good, but they can also intentionally or inadvertently do harm. There is no such thing as race-neutral algorithms, since this technology reflects the biases of those who create them and the data used in their processes. While use of these tools typically falls under Data Analysis, we treat them as their own stage of the data life cycle because of their importance in current conversations on ethical data use. We are not experts here, so our goal is *not* to give a comprehensive review of these subjects, but rather to put these technologies in the context of what we do: inform data governance processes that ensure the ethical use of data, particularly individual-level records and linked administrative data. Below, we provide examples, resources, and a list of organizations to follow for further ideas and guidance to help you and your organization/community interrogate these tools with an equity lens.

As always, definitions and distinctions are important. For the purposes of this Toolkit, we use "algorithm" as the umbrella term but also touch on these subtopics: artificial intelligence (AI), machine learning (ML), and deep learning (DL).

Algorithms	**Artificial Intelligence (AI)**	**Machine Learning (ML)**	**Deep Learning (DL)**
A set of step-by-step instructions or rules that enables automated problem solving or task completion	A technology that combines different types of algorithms to perform tasks that typically require human intelligence (e.g., recognizing patterns, learning from data, and making decisions)	A complex combination of algorithms that can learn from data and improve performance on tasks over time with minimal human intervention	A type of ML that uses a multi-layered network of algorithms to identify complex patterns within large datasets and make decisions or predictions without specific instructions
Example: A tool that displays spikes in protest activities to identify potential moments of mobilization or new movements	*Example: A tool that analyzes canvassing conversations to understand voter concerns and motivations*	*Example: Image recognition programs to detect or classify abnormalities in medical imaging*	*Example: Predictive analytics to forecast outcomes in child welfare cases*



Graphic used with permission under CC BY-ND 4.0. Vrana, J. & Singh, R. (2020). The NDE 4.0: Key Challenges, Use Cases, and Adaption.

There are strategies and tools that can and should be used to ensure transparency, assess algorithmic bias, and determine the potential positive and negative consequences of applying an algorithm in practice (see Center for Public Sector AI). Key factors include clarity and governance of algorithms, community involvement in algorithmic deployment, responsive cross-sector collaboration (see California Policy Lab & University of Chicago Poverty Lab), and continuous evaluation (see New York Department of Consumer and Worker Protection). Of course, it is also essential to define and ensure privacy within the application of algorithms.

The Weight of the Cloud

The use of algorithms, in particular AI, requires exponentially more computational power than traditional analytics. This increase in power requires additional resources—electrical power, staffing, and raw materials (minerals, water, land, etc.). While considering risk versus benefit, it is important to acknowledge the material harms that occur in the process of training and maintaining algorithms, including storage. The extraction, creation, and maintenance of these technologies rely upon predatory industries.[15] For example, the process of mining cobalt is dangerous and often involves slave labor; servers require cooling, which drains essential resources (power and water) from residents; server farms are most often built in underresourced communities of color; and the magnitude of energy use has a significant negative climate impact. The potential harms that can occur in the implementation of algorithms are linked to the material harms it takes to create them. As we move forward, let us ground ourselves in the material as we work toward ethical use.

The creation, procurement, deployment, and evaluation of algorithmic tools have significant equity implications (see City of Seattle AI Policy). To ensure ethical use of this technology, it is vital that algorithms have human oversight (see City of Boston) and are explainable when being used to make decisions or take actions that impact people's lives. For example, if an individual is denied a service as a result of the output of an algorithm, the organization must be able to explain both why the service was denied *and* what actions can be taken to access the service. Further, the organization should be able to explain how to contest the algorithmic or human+algorithm decision to deny services. The following positive and problematic practices, Work in Action, and resources will help you explore strategies for ensuring racial equity in the use of algorithms, AI, and other statistical tools.

---

[15] Png, Marie-Therese. (2022). At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance. FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.

Positive & Problematic Practices: Racial Equity in the Use of Algorithms & Artificial Intelligence

POSITIVE PRACTICE	PROBLEMATIC PRACTICE
Clearly defining the relevant terms (e.g., algorithm, AI, predictive analytics) and how they are being used by the data effort.	Failing to build understanding of the algorithmic tools being applied, including what problem they aim to solve and the potential benefits and risks of their application.
Involving community partners in early conversations about the purpose of algorithms and AI to ensure alignment with established priorities.	Deploying algorithms in high-stakes decision-making (e.g., determining program eligibility and benefits) without careful discernment by community partners in the data governance process.
Procuring technical vendors that align with the data effort's values and guiding principles (e.g., include evidence of applicable "positive practices" in selection criteria). See Narrowing Technology Solutions for IDS Initiatives, 2022	Using technical vendors purely based on legacy contracts, ease of use, or cost when they do not demonstrate understanding of the practical and ethical implications of their tools.
Clearly articulating roles and responsibilities for oversight of algorithm and AI development, implementation, and evaluation (e.g., managing data governance including data protection and quality assurance).	Not providing clear, iterative, and authentic communication channels for input regarding use of algorithms and AI.
Using algorithms and AI to provide meaningful services and supports to people represented in the data.	Using algorithms and AI for increased surveillance, punitive action, monitoring, "threat" amplification via risk scores, or other uses with no clear benefit to people represented in the data.
Shifting practice from human-in-the-loop) to human-led algorithm use (i.e., humans oversee the entire process and can override an algorithm at any point).	Believing that technology alone solves social problems (i.e., tech solutionism) while neglecting the importance of people in leveraging technological tools to enact change.
Being transparent about the use of algorithms and AI in analyses, decision-making, or other outputs (e.g., describing what data drives an algorithm and how it was tested and validated, citing the use of generative AI in report writing, identifying which department oversees decisions made by automated decision-making systems).	Relying on "black box" or proprietary algorithms or AI that do not allow for transparency or replication.

Recognizing that all data are imperfect, while continually improving the quality of data feeding algorithms, AI, or other tools.	Building algorithms based on data that reflect bias, disinformation, or power imbalances (e.g., criminal records that reflect disproportionate policing of low-income communities) or have other known data integrity issues (e.g., facial recognition and biometric data).
Using an "Algorithm / AI Red Team" to pressure test new tools for potential harms, equity issues, and worst case scenarios before deployment.	Taking a "move fast and break things" approach to developing algorithms and AI when there are implications for social welfare (i.e., plan to remediate after issues are identified).
Developing clear metrics for algorithm and AI performance and regularly auditing these tools for fairness, bias, reliability, sustainability, transparency, and explainability.	Not communicating clearly about how people's data are used by algorithms and AI and the implications for individual-level privacy.
Conducting impact assessments of algorithms to examine intended and unintended consequences and disparities of their application, compare outcomes with human decision-making, and document changes based on the findings.	Having no process to challenge decisions or outputs made by algorithms, AI, or statistical tools and seek redress for any harms.
Comparing algorithmic impact results to human decision-making to evaluate both human and automated bias.	Assuming automation yields more or less biased results without interrogating past processes.

WORK IN ACTION:
Use of Algorithms & Artificial Intelligence

Examples of Use

California Policy Lab & University of Chicago Poverty Lab, Homelessness Prevention Unit (HPU)

Janey Rountree

The Los Angeles County Homelessness Prevention Unit (HPU) is a multi-agency initiative between the LA County Department of Health Services, the Department of Mental Health, and the Chief Information Office to prevent homelessness. The HPU utilizes an innovative predictive model developed by the California Policy Lab at UCLA (CPL) to identify Los Angeles residents at the highest risk of experiencing homelessness and then proactively reaches out to them to offer assistance aimed at preventing their loss of housing. Leveraging anonymized data from 11 county agencies, the model targets individuals who have recently accessed services yet are often disconnected from other homelessness prevention programs. Since 2020, the HPU has used the model to enroll hundreds of high-need participants each year in their program, which offers flexible cash assistance, customized case management, and referrals to other services like mental health care, workforce development, and legal aid. Early findings suggest that the HPU fills a critical gap by reaching underserved, diverse populations with complex service needs. An ongoing evaluation by CPL will determine its broader impact on reducing homelessness across Los Angeles.

Allegheny County Data Warehouse, Mental Health Allegheny Housing Assessment (MH-AHA)

Kathryn Collins

The Mental Health Allegheny Housing Assessment (MH-AHA) is a decision support tool that helps the Department of Human Services (DHS) prioritize admission to supportive housing services for those with persistent and severe mental illness. The predictive algorithm uses data from the Allegheny County Data Warehouse (including age and gender, but not race) to predict the likelihood of a mental health inpatient stay or frequent hospitalization during the next 12 months. The MH-AHA underwent an extensive independent equity audit that found no evidence of discrepancies in how the tool performed across different subgroups.

Since its adoption, the MH-AHA has dramatically cut down the time individuals spend on waitlists for supportive housing. Before the algorithm was implemented, all individuals were put on a waitlist and their cases were evaluated when housing became available. Now, those that are ineligible are referred to other supportive services so they do not wait on a list indefinitely for services they will never receive. Those who are eligible are much more likely to be matched with appropriate housing within 90 days. The MH-AHA has also given DHS a greater understanding of unmet need so that they can advocate for more resources. Read more about MH-AHA and the Allegheny Housing Assessment here.

Examples of Auditing, Governance, and Regulation

City of Seattle, Generative AI Policy

Jim Loter

Seattle IT is the unified technology strategy and support department for the City of Seattle. IT processes all acquisitions of technology products, services, and data for City departments after evaluating them for suitability on the basis of a number of factors, including security, supportability, interoperability, and compliance with relevant policies. When generative AI emerged as a popular consumer-facing service in late 2022, IT recognized that it could present new risks to the City's integrity and public trust if not used responsibly. The existing IT acquisition framework was not designed to evaluate technology tools with the ability to produce substantive artificial content of unknown origin that could then be used as part of official City communications, documentation, or records.

In April 2023, Seattle IT issued a provisional policy to address the risks of using generative AI technology by City employees to conduct City business. The policy defined generative AI as "a class of AI systems that are capable of generating content, such as text, images, video, or audio, based on a set of input data rather than simply analyzing or acting on existing data." In October 2023, Seattle IT, in collaboration with the University of Washington, Allen Institute on Artificial Intelligence, and the City's Community Technology Advisory Board, released a formal policy for City employees with six fundamental principles:

- Acquisition of AI products or services, even if free, must occur via Seattle IT's process
- Employees must ensure that no copyrighted material is inappropriately published

- All generative AI produced content must be labeled
- Employees must review content to reduce bias and potential harmful or offensive material
- Employees must not submit sensitive or confidential data to AI systems
- Employees must retain generative AI inputs and outputs if they constitute public records

Seattle IT continues to work in partnership with City departments and employees to appropriately respond to growing use and application of AI.

New York Department of Consumer and Worker Protection, Local Law 144

New York City's Department of Consumer and Worker Protection has recently recommitted to enforcing Local Law 144, prohibiting the use of automated employment decision tools (AEDT) unless job candidates are informed and the tools are audited for bias. AEDTs are algorithm-based tools that utilize artificial intelligence to help make employment screening and hiring decisions with limited human input. Predictions often include the likelihood of success for a candidate in a specific role based on written application materials. New York requires that audits must be conducted by an independent auditor and include assessment of scoring methods by gender, race, ethnicity, and other intersectional categories. Local Law 144 centers transparency and assessment of disparate impact, both of which are crucial to AI governance.

City of Boston, Guidelines for the Use of Generative AI

Santiago Garces & Alejandro Jimenez Jaramillo

The City of Boston has published guidelines for the use of generative AI, highlighting both potential risks and benefits for public agency users. Boston's guidelines include sample use cases and principles for "responsible experimentation" with generative AI. The guidelines also include specific do's and don'ts so that products created using generative AI are useful, reliable, and transparent. There are also an array of resources provided, so potential users can learn more about AI tools available before adopting. Above all, the guidelines stress the importance of privacy and of ensuring that confidential information is never shared within generative AI programs. Boston's guidelines acknowledge that generative AI can be a useful tool and encourage staff to be informed consumers as they experiment, without abdicating responsibility for the outcomes.

Center for Public Sector AI (CPSAI), Public Sector Advisory Board

Justin Brown, Cassandra Madison, & Kristen Tillett

The Center for Public Sector AI (CPSAI) is a nonpartisan, nonprofit organization dedicated to bridging the knowledge gap between emerging technologies and government. CPSAI convenes government, nonprofit, independent, and private sector leaders together from across the country to equip government leaders with the knowledge, tools, and resources they need to make informed decisions about how to responsibly deploy AI and other emerging technologies in the benefits delivery space. CPSAI has a Public Sector Advisory Board composed of sitting Health and Human Services leaders to ensure that the organization's work reflects the needs of governmental leaders and builds state capacity for improved service delivery while mitigating potential harms. By creating a dedicated space for these leaders to collaborate, share insights, and collectively navigate the complexities of emerging technologies, CPSAI is working to encourage safe exploration and responsible adoption.

Resources and Tools: Algorithms & Artificial Intelligence

AI Governance Lab & Navigating Demographic Measurement for Fairness and Equity, 2024, Center for Democracy and Technology

AI and Your Agency Data Systems, 2024, Data Integration Support Center

Artificial Intelligence Framework for an Equitable World, 2024, NTEN

Principles for Accountable Algorithms and a Social Impact Statement for Algorithms, FAT/ML

Ethics & Algorithms Toolkit, 2018, Anderson, Bonaguro, McKinney, Nicklin, & Wiseman

The Data Nutrition Project, 2024, The Data Nutrition Project

First, Do No Harm: Ethical Guidance for Applying Predictive Tools within Human Services, 2017, MetroLab

Data Science and Public Policy: Tools and Guides, Carnegie Mellon University

Blueprint for AI Bill of Rights, 2022, White House Office of Science and Technology Policy

DataedX Group Resources, 2024, varied authors

Ongoing research: Data and Society, 2024, varied authors

Brandeis Marshall on Medium, 2024, Brandeis Marshall

Organizations to Watch: Use of Algorithms & Artificial Intelligence

Center for Public Sector AI

City of San Jose Government AI Coalition

Algorithmic Justice League

Black Tech Street

AI Blindspot

Racial Equity in Reporting & Dissemination

Reporting and dissemination refers to the process of communicating the findings of a project. This generally involves creating reports, presentations, visualizations, websites, social media content, artistic displays, or other products that summarize the data and findings, as well as the strategic use of these products to engage different audiences. Centering racial equity requires designing products that are engaging and meaningful to the people represented in the data, not just other data practitioners. Although a static report can convey a lot of information in a simple format, more creative methods of reporting tend to engage a wider community.

To center racial equity, first consider whom you are speaking to—your intended (and unintended) **audience**—and consider whether the design and approach fits their preferred communication style. It's important to develop a range of products with tailored messaging for these audiences. Pay attention to the **readability and accessibility** of any communication format as well to ensure not only broad appeal but also engagement with the content. For example, a presentation geared toward the public should follow basic accessibility guidelines, use person-centered language, and avoid jargon that may be otherwise appropriate for internal program staff or academic audiences. To ensure that dissemination is accessible and readable, consider using a checklist for plain language, test readability, and seek out professional translation of materials into languages relevant to your audience/community.

Beyond accessibility, it is important to be intentional about the **framing** of data and findings, which includes language, visualizations, and dissemination methods. Surveilled communities are often portrayed from a deficit perspective that reinforces harmful narratives. Using asset-based framing and intentionally choosing language that explicitly names the impacts of systemic racism, as opposed to individual or community failings, will help build trust with communities and avoid further harm. Deliberate choices can go a long way toward shifting narratives, even if they seem simple in practice. For example, when visualizing data, choose to map negative outcomes in a neutral color, rather than red. Doing so can decrease stigmatization and avoid reinforcing negative perceptions of neighborhoods or groups. Likewise, describe barriers and inequities as human-made, using active rather than passive language (see Atlanta Wealth Building Initiative in the Work in Action section).

Data and findings should be shared in a **timely manner** with communities who may be impacted by the results and those who can take meaningful action, which are not separate but rather overlapping groups. Thoughtful community-facing dissemination can facilitate change with and across different audiences. It can both encourage individual reflection and open up dialogue about systemic issues (see The Folded Map Project). It can also help build bridges to ensure that findings are trusted and used by those with the most to gain (see Camden Coalition).

Strong dissemination should **empower members of the public** to dig into the data, ensure accountability, and bring important issues to light. This requires clear documentation of how an analysis was conducted and whom to contact with questions. Allow members of the public to see themselves in the work regardless of their education or experience, by recognizing the value of all contributions through properly citing the work of others whose methods or ideas you built upon and giving attribution to all project contributors, especially community partners.

Though reporting and dissemination is the final stage in the data life cycle, the work doesn't end here. Sharing insights from the data often serves as a launchpad for new inquiries and creates opportunities to collaborate with communities to make meaning, shape future directions, and take action to improve lives.

Positive and Problematic Practices: Racial Equity in Reporting & Dissemination

POSITIVE PRACTICE	PROBLEMATIC PRACTICE
Creating a variety of products that communicate findings to different audiences through a range of formats (e.g., static and interactive, digital and analog).	Producing one output that is inaccessible to general audiences (e.g., a lengthy PDF report, publication behind a paywall).
Using checklists to review products prior to release for accessibility, cultural sensitivity, nonstigmatizing language and visuals, usability, and data integrity.	Disregarding how the findings and their presentation may impact individuals or communities (e.g., releasing data that stigmatizes student subgroups, choosing graphic color palettes inaccessible to the colorblind).
Developing tailored messaging for different audiences that considers needs and preferences, which may include the appropriate level of detail and technical jargon, reading level, language, length, and format.	Creating data visualizations or other products that are difficult to read, interpret, or make meaning of for the people represented by the data.
Disseminating information that aims to improve the lives of those represented in the data rather than bring punitive action (e.g., analyzing food purchase data to identify food deserts and guide development of grocery stores vs. to remove recipients from public benefits).	Publishing data about deficits or what's not working without including the underlying social context and suggestions for policy and practice improvement.
Providing public access to aggregate data where appropriate (e.g., dashboards, routine reports, interactive maps).	Failing to include a clear description of the underlying data used and necessary context for interpretation.
Including stories as a complement to quantitative findings in order to better contextualize the lived experience represented by the numbers.	Attempting to describe individual experiences with aggregate or "whole population" data without examining disparate impact based on race, gender, and other intersections of identity.
Providing clear documentation of the data analysis process along with analytic files to ensure replicability and reproducibility of results.	Making documentation indecipherable to those who do not regularly work with the data and not including contact information for those who have questions.
Giving proper credit to all individuals and groups that supported the project, including co-authors, community partners, data providers, funders, reviewers, and participants.	Not giving attribution to the work or ideas of others upon which the project builds.
Conducting impact analyses throughout the project to assess: how does this work mitigate, worsen, or ignore existing disparities?	Failing to respond to impact analyses, community feedback, data errors, or harms identified in reporting.

WORK IN ACTION:
Racial Equity in Reporting & Dissemination

The Atlanta Wealth Building Initiative, Building a Beloved Economy: A Baseline and Framework for Building Black Wealth in Atlanta

Alex Camardelle & Jarryd Bethea

The Atlanta Wealth Building Initiative uses administrative data to highlight structural obstacles to reducing the massive racial wealth divide in Atlanta, Georgia, as well as possibilities for progress. Their report, Building a Beloved Economy: A Baseline and Framework for Building Black Wealth in Atlanta, features the structural determinants of a Black wealth framework—a visual depiction of resources as a flowing river able to be obstructed by human-made forces that act as a dam. The visual communicates the importance of a well-fed, nurturing environment "that supports the development of policies and programs that promote the complete freedom of Black people to grow and maintain wealth." The report also offers 18 potential policy solutions that empower any partner—from community organizers and researchers to policy makers and investors—to advance solutions. Though the audience is broad, the messaging is tailored to communicate the deep roots of the issue and to empower readers toward action in relationship with those affected. Administrative data are clearly used for the explicit purpose of improving the lives of and opportunities available to those represented in the data. Read the full report here.

Camden Coalition, Youth Ambassadors Program

Taylor Brown & Martiza Gomez

In the wake of the COVID-19 pandemic, as fear and misinformation caused hesitation around testing and vaccination, the Camden Coalition created a Youth Ambassador program to build trust among residents of Camden, New Jersey. Around 60 volunteers between the ages of 13 and 22 received training to stay up-to-date on the latest information surrounding COVID-19, so they could communicate with vaccine-hesitant people in an impactful way. These Youth Ambassadors also provided insight on how to reach other young people. For example, vaccine outreach was conducted at in-person events and through canvassing, because many young people were skeptical of information on social media. By working with members of the community whose experiences they could relate to, the youth volunteers were able to share accurate, timely information and decrease hesitancy in their communities. Learn more about the Youth Ambassador program here and read about their lessons learned. The Community Ambassadors program, which the Youth Ambassadors program was based on, continues to do public health messaging work locally. Learn more about that program here.

The Folded Map Project

Tonika Lewis Johnson

The Folded Map Project is a multimedia art project that investigates segregation in Chicago through the concept of address pairs and "map twins." Created by Tonika Lewis Johnson, the project visually contrasts side-by-side photos of homes with the same address but miles apart on the racially and economically distinct North and South sides of the city. In addition to these images, the project includes portraits of residents in front of their homes and video-recorded conversations between them and their "map twins," homeowners with the same address on the opposite side of the city. This art project not only reveals the stark realities of investment on the North Side and disinvestment on the South Side, but also highlights the power of personal conversations to deepen understanding of these systemic issues. It invites viewers to reflect on how the social, racial, and institutional forces driving segregation impact us all. Explore the project here: Folded Map Project. You can also check out the Folded Map Action Kit, which offers concrete steps for individuals to take to address segregation.

Interactive Arts Exhibit

In a large county in the southeastern U.S., a collaborative effort serving children and youth partnered with a local artist to curate an interactive arts exhibit based on a Community Participatory Action Research (CPAR) project they conducted in 2023. The CPAR project focused on understanding the impacts of an involuntary admission to a hospital for a mental health crisis. The CPAR project brought together as co-researchers youth and parents of children who had experienced an involuntary admission and mental health system professionals. In 2024, research findings were shared with a group of artists (both commissioned and selected through an open call) and translated into an interactive arts exhibit. Opening night of the exhibit included live performances with the display of visual art pieces, as well as art-making opportunities for attendees. Follow-up events throughout the month included meetings with elected officials and system leaders and community conversations with artists and system professionals. The arts exhibit was funded with support from national and local donors. To date, the exhibit has reached hundreds of constituents to shift the conversation around the impacts of this policy. Most importantly, youth, parents, and families who participated in the exhibit shared how the experience reduced the stigma associated with mental illness, helped them feel less alone and isolated, and educated them on opportunities for improving the experience and outcomes when mental health crises occur.

Resources: Reporting and Dissemination

Accessible Data Viz Is Better Data Viz, 2018, Storytelling with Data

Connecticut Data Visualization and Accessibility Guidelines, 2023, CT Office of Policy & Management

Do No Harm Project, 2021, The Urban Institute

The Equitable Communications Guide, 2023, The Innovation Network

Understanding Data Accessibility for People with Intellectual and Developmental Disabilities, 2021, ATLAS Institute

Working with Tribal and Indigenous Data Across the Data Life Cycle

One key concept that was missing in our last Toolkit, and which applies to work at every stage in the data life cycle, is Tribal data sovereignty. **Tribal Nations have an inherent right to govern their own data and must be consulted about any data that concerns them.** Tribes are sovereign nations, meaning that they have a government-to-government relationship with state and federal agencies and hold the authority to make decisions about their people and their lands—referred to as "Tribal sovereignty." This includes decisions about the collection, ownership, and use of their data, which may be referred to as "Tribal data sovereignty" or "Indigenous data sovereignty" when speaking more broadly than the U.S. Tribal Nation context. In some cases, Tribes may also have legal authority to access data collected by other governments.

It is critical to engage in ongoing Tribal Consultation whenever Tribal data use is considered. Every Tribal Nation is unique and will have their own protocol for Tribal Consultation. Some may also have their own data protection policies, research codes, and Tribal Institutional Review Boards (IRBs) that must be included in the process. These are especially important to abide by given the long history of Tribal Nations and Tribal members being exploited in research.[16] For more guidance on principles and best practices for Tribal Consultation, see and reach out to your agency or state Tribal Affairs Office or Tribal Liaison. For more guidance on principles and best practices for Tribal Consultation, see Guidance and Responsibilities for Effective Tribal Consultation, Communication, and Engagement and reach out to your agency or state's Tribal Affairs Office or Tribal Liaison.

> **Tribal Consultation:** A formal process of engaging meaningfully with Tribal partners about their data (or other matters concerning their people or lands)

When working with any data, understand the Native population that may be represented in your data and use language that is specific to that context. For example, the term "Indigenous" broadly refers to peoples with pre-existing sovereignty who were living together as a community prior to contact with settler populations and may be appropriate to use as an inclusive term for Native peoples around the world. However, if you are talking about data for a specific Tribal Nation, use the name of that nation (e.g., Cherokee Nation, Lakota Sioux) unless directed otherwise by a Tribal partner. The United States alone has 574 federally recognized Tribes, many others that may be recognized only by states, and still more Tribal Nations that are not officially recognized.

The terms "American Indian" and "Alaska Native" broadly refer to the Tribal Nations across the United States and may be appropriate to use when working with data that covers many Tribes within this geographic context. In Canada, the equivalent term is First Nations, Inuit, and Metis. Keep in mind, however, that every Tribe has its enrollment policy—not everyone who identifies as Indigenous may

---

[16] Saunkeah, B., Beans, J. A., Peercy, M. T., Hiratsuka, V. Y., & Spicer, P. (2021). Extending research protections to tribal communities. *American Journal of Bioethics*, *21*(10), 5-12.

be enrolled in a Tribe. Additionally, urban Tribal members may not be given the same protections from research as those living on Tribal lands because of living outside of the geographic borders in which their Tribal Nation holds governance authority. Nevertheless, researchers should familiarize themselves with appropriate research approval processes for the urban Indigenous communities with whom they wish to partner.[17]

Although language may shift over time and not all terms have universal definitions, the following resources provide an introduction to Native identities and their distinctions. You can also rely on Tribal partners for guidance about how they should be referenced in datasets and written materials.

- Terminology Style Guide, Native Governance Center, 2021
- Indigenous Identity: More Than "Something Else," Native Americans in Philanthropy, 2020
- The United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), Indian Law Resource Center, 2007

After completing both informal and formal Tribal consultation processes to engage meaningfully with Tribal and other Indigenous partners around data use, you must then apply an equity lens at each stage of the data life cycle as it relates to this population.

- When **planning** a project, develop a data governance framework with Tribal or Indigenous partners that includes a plan for stewarding their data (e.g., Tribal affiliation, geographic boundaries of Tribal Jurisdictional Areas or Ancestral Lands). Tribal counsel should be at the table when legal decisions are made around access and use. If there is ambiguity about whether a particular use is permissible, it is important to prioritize getting social license and approval from Tribal leaders and counsel. Also be cognizant that your priorities may not be a priority for Tribal or Indigenous partners and you may need to shift directions in order to come to agreement about project scope and aims.
- When designing **data collection,** work with Tribal partners to discern how to gather data that can best identify Tribal members, lands, resources, and services and allow for meaningful analysis (e.g., collecting Tribal affiliation data by including a comprehensive list of options to pick from while also allowing space for people to self-identify). Ask Tribal partners how their community would like to participate in data collection.
- To honor the government-to-government relationship that Tribal Nations are entitled to, make sure that there is a clear plan in place to grant Tribal or Indigenous partners **data access** and to return any data and analyses to these communities in ways that are helpful to them.
- When doing demographic **data analysis,** consider counting people who identify as Indigenous alone *or* in combination with other identities (e.g., checking a single box for American Indian/Alaska Native (AI/AN) vs. checking multiple boxes that include AI/AN) and people who have written in their affiliation(s) when the appropriate box was not listed on a data collection tool.

---

[17] Haozous, E. A., Lee, J., & Soto, C. (2021). Urban American Indian and Alaska native data sovereignty: ethical issues. *American Indian and Alaska native mental health research*, *28*(2), 77.

- Contextualize the results of **data analysis** with the history of colonization and the conditions it created for Tribal or Indigenous communities—especially when interpreting and presenting results related to challenges.[18]
- When considering the use of **algorithms or AI,** look to frameworks for ethical use that are specific to Native or Indigenous people rather than relying on generic guidance.[19]
- Involve Indigenous people in the design of **reporting and dissemination,** tailoring products to methods and practices of information sharing within the Tribal community (see examples from the Kahnawake Schools Diabetes Prevention Program and Te Kāhui Raraunga Charitable Trust).

WORK IN ACTION:
Working with Tribal and Indigenous Data

Minnesota's Missing and Murdered Indigenous Relatives (MMIR) Task Force Report

Nicole MartinRogers

American Indian women and girls make up less than one percent of Minnesota's overall population, but they account for eight percent of all murdered women and girls in Minnesota (2010-2018). Each month, between 27 and 54 American Indian women and girls are reported missing (2012-2020). Moreover, when someone goes missing or dies under suspicious circumstances, the victims and their relatives are less likely to experience justice. The overrepresentation of American Indians among missing persons and these disproportionate justice outcomes are due to a combination of historical factors that started when Turtle Island (North America) was colonized by Europeans. Since that time, Indigenous communities have experienced genocide and traumas such as boarding schools and the removal of Indigenous children from their families in the name of "child welfare." Indigenous peoples have been subjected to systemic policies and practices that have perpetuated poverty, family dysfunction, addiction, and the painful loss of culture.

In 2019, the Minnesota legislature recognized that American Indian people are more likely to go missing or become the victims of violence when compared to White Minnesotans. With no previous examples of legislation that sought to address violence against Indigenous women and communities in the United States, now state Senator Mary Kunesh reached out to the Indigenous community for guidance on creating a statewide Task Force. She invited nonprofits and grassroots advocacy groups from across Minnesota to convene at the State Capitol and over the phone to discuss potential legislation. Indigenous women from across Minnesota provided compelling testimony that ultimately ensured that the bill passed with unanimous bipartisan support.

---

[18] Phillips-Beck, W., Star, L., & Leggett, S. (2024). Navigating Indigenous Data Sovereignty: A Decolonizing Approach to Understanding Opioid Use Amongst First Nations in Manitoba. *International Journal of Population Data Science*.; First Nations Health and Social Secretariat of Manitoba: https://www.fnhssm.com/.

[19] Bergin, P. (2023). How Indian Tribes Can Utilize AI Technology and Legal Issues to Consider.; Crepelle, A. & Murtazashvili, I. (2023). Artificial Intelligence on Indian Reservations in the United States: Prospects and Challenges. SSRN.; Sorrentino, D. (2024). TribalNet 2024: Artificial Intelligence and Data Sovereignty Take Center Stage. FedTech.; Walter, M. & Kukutai, T. (2018). Artificial Intelligence and Indigenous Data Sovereignty. Input paper for the Horizon Scanning Project on behalf of the Australian Council of Learned Academies.

The bill created the Missing and Murdered Indigenous Relatives (MMIR) Task Force and provided significant resources to conduct background research and facilitate discussions. Task force members engaged advocacy organizations, as well as representatives from the Bureau of Criminal Apprehension, the Peace Officer Standards and Training board (POST), and the Minnesota departments of Health and Human Services. Across these groups and at the highest levels of state government, there was universal commitment to share information (as legally allowable) and to brainstorm solutions.

The Task Force set out to acknowledge the root causes of MMIR injustice and highlight the policies, practices, beliefs, and relationships that affect families and communities during missing persons and murder investigations; demonstrate that subgroups of Indigenous people are at particularly high risk; and show the ways specific industries contribute to the harm of Indigenous women. The report also acknowledges the jurisdictional complexities that contribute to MMIR injustice and used Task Force research to recommend concrete and practical solutions, such as Tribal Community Response Plans, so that these issues would no longer be an excuse.

The Minnesota Department of Public Safety, in conjunction with the Task Force, contracted Wilder Research to assist in conducting research and writing their final report to the legislature. Wilder Research conducted a literature review, a federal and state policy review, and nearly 30 interviews with experts from state agencies and advocacy groups to identify key opportunities and challenges related to MMIR injustice in Minnesota. These experts were also invited to speak to the Task Force, including a leader of the Canadian Missing and Murdered Indigenous Women project, and the head of the Minnesota Missing Persons Clearinghouse. During the summer of 2020, Wilder Research hosted a series of virtual Task Force sessions to review key findings from their research to inform and refine proposed policy mandates. Using all of the findings from the previous year, the Task Force and their partners submitted a final report to the Minnesota legislature in 2020.

Since the report was submitted, the MMIR office, established in 2021, and Wilder Research have been working to eliminate MMIR injustice in Minnesota. Senator Kunesh and colleagues passed legislation that created a reward fund to help solve MMIR cases, and this reward fund has an advisory board that includes the former Task Force chair, along with family members of MMIR victims. The MMIR Office has also created a community advisory board that provides overall direction and oversight for its work. Now, the MMIR Office is creating a dashboard that combines indicators from the final report—from several state data systems and other sources—to give a clearer picture of the issue and trends in Minnesota. This public dashboard will give the MMIR Office, advocates, agencies, and families shared information on MMIR cases, and will also provide more accountability regarding whether the efforts are having the desired impacts.

Going forward, the MMIR office is working with POST to update training requirements and Peace Officer testing to ensure that the process for filing missing persons reports is clear and followed by officers, especially when there are multiple jurisdictions involved. The office is also encouraging Indigenous Nations to work with local, state, and federal law enforcement to create Tribal Community Response Plans, which can be deployed when an Indigenous person is victimized or goes missing. All of these policies are designed to center Tribal sovereignty and empower Indigenous Nations to protect their communities and create justice.

Elements vital to success: Qualitative data—it is necessary to pair with quantitative methods to truly understand a complex issue like the MMIR injustice.

What we wish we had been told: Just because data are interesting and relevant doesn't mean they need to be part of an integrated data system.

Lessons Learned:

- No one person or agency knows about all of the data that could or should be included in the analysis of a complex issue like the MMIR injustice; multiple sources of data and perspectives from key experts representing various components of the issue are needed to increase understanding of the issue.
- Data can't be used to simply admire a problem unless a concerted effort is made to present concrete findings that directly lead to actionable recommendations for change.
- Centering the experiences of community members and people with lived experience can help to focus on areas with the potential for greatest impact.

Resources: Tribal Data Sovereignty

Advancing American Indian & Alaska Native Equity, 2022, Evans-Lomayesva, Lee, & Brumfield

CARE Principles for Indigenous Data Governance, 2020, Russo Carroll, Garba, et al.

First Nation's Government Information Center, 2024, The First Nations Information Governance Centre

NCAI Resolution, 2024, National Congress of American Indians

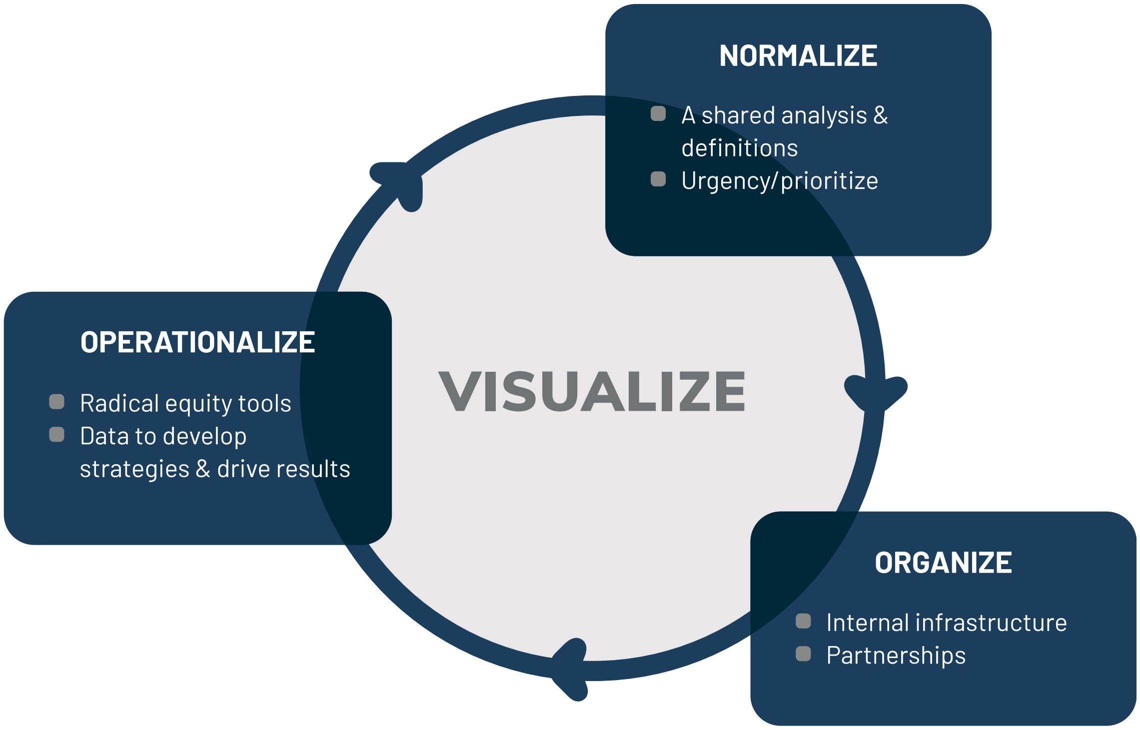
Operationalizing the CARE and FAIR Principles for Indigenous Data Futures, 2021, Carroll, Hudson, et al.

Tribal Data Governance and Informational Privacy: Constructing "Indigenous Data Sovereignty," 2019, Tsosie

What's Next?

Wherever you are, build from where you have been. We recommend using the framework outlined by the Government Alliance for Racial Equity (GARE) to **visualize, normalize, organize, and operationalize** racial equity throughout data integration.

We urge you to begin by reflecting on your individual and institutional starting point for centering racial equity within data integration. Next, use the core questions in each of the Visualize, Normalize, Organize, and Operationalize sections to reflect on ways that you and your organization can continue to grow in these areas. The links included in each section and Companion Workbook activities will help you further assess and take action toward centering equity. It is important to note that this work is nonlinear and iterative.



NORMALIZE

- A shared analysis & definitions
- Urgency/prioritize

OPERATIONALIZE

- Radical equity tools
- Data to develop strategies & drive results

VISUALIZE

ORGANIZE

- Internal infrastructure
- Partnerships

Image used with permission. Source: GARE Communications Guide, May 2018

Visualize

How can data be used to share power, build relationships, and deepen trust? What is your shared vision for how data can be used to center racial equity? What are your guiding principles for working toward this vision?

See Companion Workbook, 1.IV, Co-create a vision and guiding principles for centering racial equity throughout data integration.

Normalize

How has your lead agency/collaborative acknowledged the importance of a racial equity lens and demonstrated a commitment to engage in data integration efforts that are legal, are ethical, and center equity?

Assess your current activities using these resources:

- Community Engagement Assessment Tool, Nexus Community Partners, 2020
- Data Ethics Workbook, UK Department for Digital, Culture, Media & Sport, 2018
- Health Equity Impact Tool, State Health & Value Strategies, 2024
- Tool for Organizational Self-Assessment Related to Racial Equity, Coalition of Communities of Color, 2014

Organize

In your site context, how will the community and government learn, work, and be mutually accountable for using integrated data to inform, evaluate, and co-create structures, policies, practices, and narratives for equity?

Questions to guide your thinking should include these:

- How will community expertise be at the forefront throughout the data life cycle? What capacity will need to be developed to ensure that this occurs?
- How will data use help communities interrogate systems, rather than just inform how to "treat" communities with additional services and programs?
- How will a racial equity lens be incorporated throughout the data life cycle?
- How will the culture, policies, practices, and expectations of the agency/collaborative shift to center racial equity?
- How will the agency/collaborative initially focus upon, continuously learn from, and sustain institutional-systems change?

Operationalize

What approaches will be most effective for integrated data infrastructure development and data use? Consider each question across three core audiences: policy makers, agency, and community. How can practices such as capacity-building, collaborating and power sharing, and centering and empowering community be designed to create impact?

Questions to guide your thinking should include these:

- What work has been done with policy leaders, agency directors, department staff, and front-line employees to prepare them for this work?
- How has the agency/collaborative engaged with community members, and in what manner?

- How will the agency/collaborative ensure that the benefits of data integration outweigh the risks?
- How will community members whose data are in the system receive capacity-building? What structures will be implemented, including funding, training, and ongoing relationship building?
- What role does community organizing play in the design and use of data?
- What guides accountability and power sharing between community members, policy makers, agency staff?

We recommend these resources to move forward:

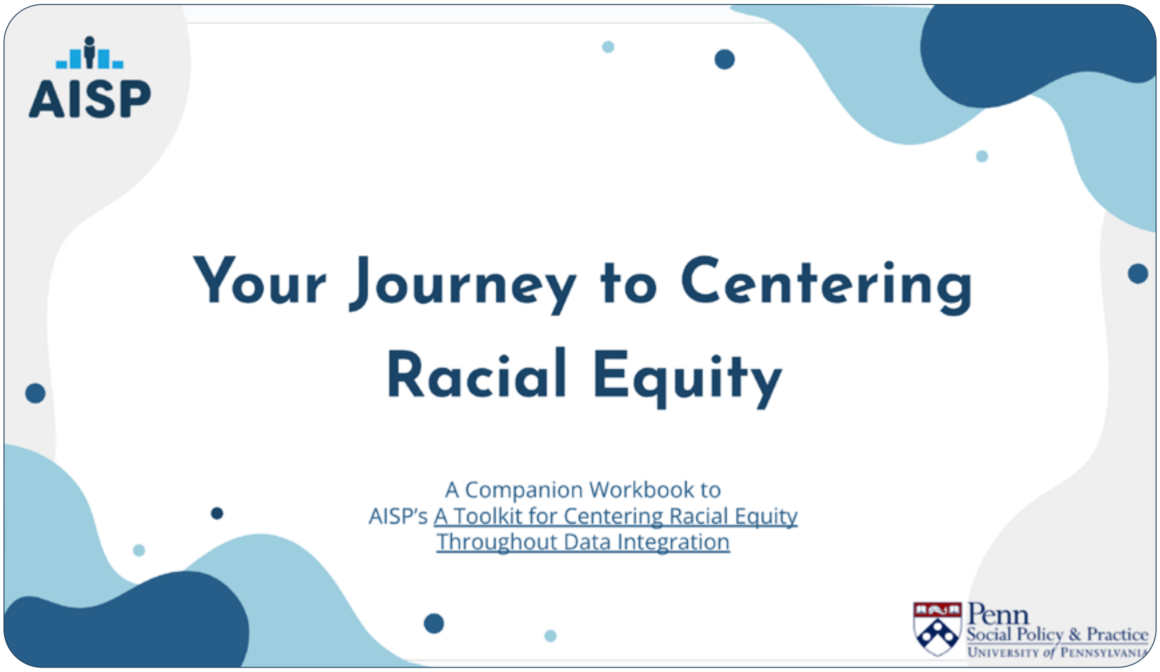
- Advancing Racial Equity: A Framework for Federal Agencies, 2022, Government Alliance on Race & Equity (GARE)
- Awake to Woke to Work, 2018, Equity in the Center, a Project of ProInspire
- Racial Equity Data Road Map, 2020, Commonwealth of Massachusetts
- Racial Equity Toolkit: An Opportunity to Operationalize Equity, 2015, Government Alliance on Race & Equity (GARE)

Our Work Together

We end this Toolkit at the beginning. This work is big, messy, and can feel glacial. All you can do is get started.

While building and maintaining data infrastructure is an inevitable and essential project of our time that allows agencies to share and use data in new ways, racial equity and community voice have rarely been centered in these efforts. **We envision data integration and use as a means to confront racism, expose injustice, act on our shared values, and elevate lived experience.** We aim to use collective knowledge and skills to advance government transparency and accountability in data use, which is critical to building trust, community well-being, and improved outcomes.

We encourage you to use the resources, stories, and Companion Workbook shared here to shift your awareness and practice toward actualizing a data infrastructure that centers racial equity. We believe our work together can support sustained investment in government and community capacity for collaboration around ethical data use.



AISP

Your Journey to Centering Racial Equity

A Companion Workbook to
AISP's A Toolkit for Centering Racial Equity Throughout Data Integration

Penn Social Policy & Practice
University of Pennsylvania

We encourage you to share your experiences in doing the important work of centering racial equity throughout the data life cycle. Contact us at www.aisp.upenn.edu, so we can learn alongside agencies and continue to share best practices and Work in Action.

Terms

Data Sharing: The practice of providing access to information not otherwise available.

Data Integration: The process of bringing together data from different sources, which often includes identifiable information (e.g., name, date of birth, SSN) so that records can be linked at the individual level.

Administrative Data: Data collected during the routine process of administering programs.

Administrative Data Reuse: Using administrative data in a way not originally intended, e.g., for research.

Asset-Framing: defining people and communities by their contributions and strengths before noting challenges and deficits.

Bias: The tendency to favor one perspective, outcome, or group over others, often in an unfair or unbalanced way. Bias can stem from personal beliefs, cultural influences, or systemic factors, and leads to distorted judgment or decision-making. Bias may be conscious (explicit) or unconscious (implicit) and affects all stages of the data life cycle. Common examples of bias in data include sampling bias and confirmation bias.

Community: A group of people who share a common place, experience, interest, or a larger system that people are a part of (e.g., youth in foster care).

Community Engagement: The process of working collaboratively with and through groups of people to address issues affecting the group's well-being. Community engagement should include authentic processes at all stages of a project. Centering the community in agency work is necessary to achieve long-term and sustainable outcomes.

Consent: Explicit permission regarding the collection, storage, management, and use of personal information. Individuals can give active (i.e., opt-in) or passive consent (i.e., implicit or opt-out). Consent must be freely given, specific, informed, and unambiguous. See Future of Privacy Forum & AISP.

Data: Information collected to help decision-making.

Data Ethics: A branch of ethics that evaluates data practices with the potential to adversely impact people and society. Ethical concerns should be considered and addressed at all stages of the data life cycle. See Open Data Institute.

Data Governance: The people, policies, and procedures that determine how data are used and protected.

Data Infrastructure: The systems, technologies, and processes for using, storing, securing, and interpreting data. This includes hardware, software, and organizational practices.

Data Minimization: The principle of limiting or minimizing the collection, storage, and disclosure of data to only what is necessary to accomplish a specific use. Data minimization is an important principle that supports privacy and ethical data use.

Data Privacy: A fundamental right, guaranteed by law, that provides individuals control over the collection, use, and dissemination of their data. Agencies must put procedures in place that protect how personal information flows, protect the freedom of thought and exploration, and protect one's dignity and reputation. See Future of Privacy Forum & AISP.

Disaggregation: Breaking down large data categories into more specific subcategories.

Equity: The elimination of privilege, oppression, disparities, and disadvantage to co-create a just, fair, and inclusive society in which all can participate, prosper, and reach their full potential. See PolicyLink.

Racial Equity: A process of eliminating racial disparities and improving outcomes for everyone. It is the intentional and continual practice of changing policies, practices, systems, and structures by prioritizing measurable change in the lives of people of color. See Race Forward.

Racism:

- **Individual Racism:** The beliefs, attitudes, and actions of individuals that support or perpetuate racism in conscious and unconscious ways.
- **Institutional Racism:** The ways in which policies, procedures, and practices of parts of systems (e.g., schools, courts, transportation authorities) or organizations influence different outcomes for different racial groups.
- **Structural Racism:** The normalization and legitimization of an array of dynamics—historical, cultural, institutional, and interpersonal—that routinely advantage Whites while producing cumulative and chronic adverse outcomes for people of color. See Equity in the Center.

Social Justice: The proactive reinforcement of policies, practices, attitudes, and actions that produce equitable power, access, opportunities, treatment, impacts, and outcomes for all. See Equity in the Center.

AISP

Actionable Intelligence for Social Policy

University of Pennsylvania

3701 Locust Walk, Philadelphia, PA 19104

www.aisp.upenn.edu