

PRINCIPLES FOR ADVANCING EQUITABLE DATA PRACTICE

MARCUS GADDY AND KASSIE SCOTT

JUNE 2020

Research organizations, governments, philanthropy, and nonprofits have expressed strong interest in using data to support the well-being of communities. But too few people recognize that data are not neutral. The decisions people make about which data matter, what means and methods to use to collect them, and how to analyze and share them are important but silent factors that reflect the interests, assumptions, and biases of the people involved. We argue for an intentional focus on thinking critically about the creation and use of data. We believe data can be a tool to advance equity and well-being, but achieving that requires an explicit recognition that those factors affect the validity and legitimacy of data-informed decisions.

Many people, especially people of color and people with disabilities, have faced harm when powerful interests have misused and abused their data. For example, policymakers have used [voting data](#) to craft laws to disenfranchise Black voters, and a major college admissions exam company faces [litigation](#) after being accused of flagging students' disability status in data sold to colleges and universities. Moreover, long-standing barriers to education, wealth, and political power have marginalized many people in data-informed decisionmaking. Technical experts and political leaders do not always value lived experience and at times have minimized the knowledge that community members hold—even when it may [shed new light](#) on pressing issues.

One way to promote data for good is to broaden the conversation about incorporating an equity lens to everyone who works with data. Organization and agency leaders and [data stewards](#), managers, and users all have

a duty to manage and use data responsibly. Researchers have been thinking about aspects of equity for some time because of institutional review boards (IRBs) that have oversight authority to protect people involved in formal research. However, even at their most effective, IRBs focus narrowly on direct individual harm and may prioritize institutional interests over protections for communities or groups of people, even if inadvertently. In addition, a great deal of data collection and analysis occurs outside universities and large research institutions, in settings without IRBs. We need a new way of thinking about the relationship between the use of data and the people and places being studied.

To elevate that discussion, this document uses the seminal [Belmont Report's](#) three principles for protecting human subjects (which are followed by IRBs) to frame ways to make affected communities and groups of people a first-tier consideration throughout the data life cycle—from the first plans to collect data to the time they are destroyed. These principles can be applied to administrative data collected in the operation of government agencies, nonprofits, or businesses; data and derived statistics from surveys; or qualitative data gathered from focus groups or interviews. The examples presented in this document are not exhaustive. Instead, they show that a community-centered approach to data is possible across sectors and types of organizations and in the different ways that people interact with data. We see this approach as the beginning of a conversation about creating a less harmful and more just data environment. We hope that this document promotes common values across sectors and inspires new thinking about how to put these principles into practice.

PRINCIPLES

Principles are enduring. They matter because they can guide actions in a wide variety of contexts. In the collection and use of data, that adaptability is essential because technology and innovation are constantly changing the boundaries of what is possible—both for good and bad.

This document relies on the principles outlined in the Belmont Report, a foundational work that has guided human subjects' protections for more than 40 years. Although the Belmont Report was originally created to guide research, its three principles—beneficence, respect for persons, and justice—inform fields as diverse as nursing and information technology. Building on that foundation, this document extends the principles to data practice.

Beneficence is the commitment to maximize benefits and avoid causing harm to the extent possible, even if it is not a formal or legal requirement.

Beneficence centers the importance of considering risks and benefits holistically. Even large benefits do not always outweigh risks, particularly when risks are great and the people who bear the risks may not directly benefit.

Respect for persons is the responsibility to uphold people's power to make decisions that are in their best interest and to protect people who do not have that power.

People can make informed decisions when they have information, the capability to understand it, and the freedom to act on it. When age, disability, or other circumstances, such as language or literacy, limit any of those three elements, people deserve special consideration and protection.

Justice is the commitment to the fair distribution of burdens and benefits among people.

This principle underscores that communities and people who bear risks and burdens deserve to benefit. It also emphasizes the responsibility to ensure that undue burden is not put on people unlikely to benefit.

PRINCIPLE-ALIGNED PRACTICES FOR THE DATA LIFE CYCLE

Different decisionmaking points arise throughout the data life cycle that inform the intent, use, form, and control of the data. The data life cycle has four stages—acquisition, processing and analysis, dissemination, and disposition—and for each of them, we provide examples of ways to infuse the Belmont Report's principles. These examples capture only a few of the possibilities, but they should illustrate how everyone who collects, views, or uses data can consider these principles in their work.

ACQUISITION

Acquisition is the stage at which people decide which data to collect and why (conception), determine how to collect them (instrumentation), and take action to obtain them (collection). Even if you acquire only secondary data, or data someone else collected, it is important to be aware of the primary data collectors' motivations—why they wanted the data and how they collected them—before you plan and conduct your analysis.

CONCEPTION

Conception is when people form ideas about what the data are meant to show. This is when people frame data around “what” (e.g., what is health?) and “who” (e.g., who is healthy?).

Justice

Seek and include communities' interests in design considerations.

Deciding what data to collect largely centers on the intended purpose. For example, when the scope of a project is inherently about the funder's interests, community-level interests and needs are often not an intentional design consideration. In turn, the data that come from the community often end up being of little use to the community. Holding listening sessions with community residents, speaking with community leaders and activists, and engaging local service nonprofits and agencies are concrete ways to learn what data the community thinks are relevant to improve their lives.




Resource: [How to Add Informed Consent to Your Responsible Data Practices](#)

Beneficence

Be aware of how sensitive topics can affect people and communities.

Drug use, sexual behaviors, voting, and income are common examples of sensitive topics. Data on sensitive topics can expose individuals and even whole communities to possible harm, and a holistic evaluation of risks and benefits is crucial. The benefits should outweigh the risks. When benefits to the people or community that bears the risks are small, indirect, or unlikely, omitting sensitive topics may be appropriate. Where that is not possible, such as in some social service contexts, reframing sensitive topics using less-sensitive proxies or working with community members to establish legitimacy for the inclusion of sensitive topics in the scope of inquiry are alternative options.

 **Resource:** [Undertaking Sensitive Research in the Health and Social Sciences: Managing Boundaries, Emotions and Risks](#)


INSTRUMENTATION

Instrumentation is when people make decisions related to measurement: for example, designing surveys, creating administrative forms, and developing protocols for interviews and focus groups.

Beneficence

Minimize the amount of personally identifiable information (PII) collected.

The collection of [PII](#) such as names, Social Security numbers, or dates of birth can gravely compromise people's finances, privacy, and safety, especially when those data are attached to other records. Only asking for PII when necessary or using alternative unique identifiers, such as randomly generated unique IDs, in place of very sensitive forms of PII whenever possible can help reduce the risk to individuals and communities.

 **Resource:** [Ten Guiding Principles for Data Collection, Storage, Sharing, and Use to Ensure Security and Confidentiality](#)

Beneficence

Be conscientious about re-identification risk.

Datasets can unintentionally expose sensitive information about people when combined with other data from public or proprietary sources. Demographic information is both widely available and in sensitive datasets, and it can be used to link otherwise anonymized data. For example, race, even when few other personal characteristics are included, can provide enough information to re-identify someone in a community that is not racially diverse. When designing a data collection instrument, analysts should understand how the data could be put together to re-identify respondents and take measures to avoid it. On the back end, sometimes data managers have to remove variables to more fully anonymize datasets and mitigate risk. And at the outset, data collectors should consider whether certain information needs to be collected at all. If it must be, putting procedures in place to protect respondents is key.

 **Resource:** [Assessing and Minimizing Re-Identification Risk in Research Data Derived from Health Care Records](#)


COLLECTION

Collection is when people gather data by fielding surveys, collecting or acquiring administrative information, and recruiting for and leading interviews and focus groups, among other methods.

Justice

Avoid undue burden.


It is important to be conscientious about the labor and the mental or emotional stress it takes for people to respond, especially when they may not have a way to opt out. Some instruments, such as intake forms for services, may include many elements. Some elements are essential, such as those that are included to comply with statutory or oversight requirements, while others are not. Eliminating the collection of nonessential elements—particularly when the data are not actively used—will reduce undue burden.

 **Resource:** [Ten Guiding Principles for Data Collection, Storage, Sharing, and Use to Ensure Security and Confidentiality](#)

Respect for persons

Incorporate informed consent whenever possible, even if not formally required.

Though informed consent is common in research settings, a lot of personally identifiable data is collected in nonresearch settings for purposes that may never be disclosed. Informing people in ways they can understand of the kind of information collected about them, its purpose, and the potential risks and benefits is important. In cases where that is not feasible, which commonly happens with passive data collection like public video recording, written or verbal notices to inform people that they are under observation allow them to make more informed choices about what they say or do.

 **Resource:** [How to Add Informed Consent to Your Responsible Data Practices](#)

PROCESSING AND ANALYSIS

Processing and analysis is the stage at which people decide how to tabulate and interpret data. This includes determining which data to include or exclude in analyses, how to process the data to create new variables or measures like indexes, and which people will be involved in the process to decide what the data mean.

Beneficence

Seek out and incorporate communities' interpretation of the data.

Analysis often happens apart from the people and places being studied, which can lead to incorrect findings and can even bring harm to a community. Giving people the opportunity to offer their interpretations of the data can reduce the risk of making incorrect or harmful interpretations. A data walk is one way to engage people in conversations around data and to facilitate those conversations. Other engagement possibilities are organizing formal structures, like community-led panels and committees, and tapping into partnerships with community-led organizations, like parent associations and resident councils.

 **Resource:** [Data Walks: An Innovative Way to Share Data with Communities](#)

Beneficence

Be transparent about the limits of the data.

All data have limits, and it's important to be clear about what data do—and do not—mean. Some limits will be clear from the definitions in a dataset's codebook or made explicit in a person's statements in an interview or focus group, but analyses can still overextend the data to make points they do not support. For example, it would be a mistake to conclude that data showing higher rates of school disciplinary sanctions against students of color mean that students of color have a higher underlying incidence of disruptive behavior. A careful analysis would also examine enforcement and if no data exist would clearly acknowledge that a lack of data left key dynamics unexplored. Being transparent about what data informed an analysis and clearly documenting the process for making analytical decisions are useful ways to facilitate community members' ability to respond proactively when errors enter an analysis.

 **Resource:** [Advancing Better Outcomes for All Children: Reporting Data Using a Racial Equity Lens](#)

DISSEMINATION


Dissemination is the stage at which people publish data and/or report findings from their analysis.

Beneficence

Account for how publication may reinforce inequities or close disparities.

Data and research findings, particularly on politically sensitive subjects, can have an outsize impact on people and communities that have faced systematic marginalization and neglect. Topline findings that may be neutral statements of fact—for example, that a school district's standardized test scores have fallen—can easily be misconstrued. In this example, the finding can further a narrative of “failing schools,” fueling dynamics like white flight that reinforce school segregation and its many inequities. Considering how data findings can be misinterpreted to harm communities and mitigating that are important aspects of community-centered work. Some ways to do this are (1) proactively learning about the social and policy context of the topic of your publication; (2) paying attention to and being forthcoming


about whose interests and voices are present in—or absent from—your work; and (3) receiving critical input before publication from the people who will bear the risks of the findings.

 **Resource:** [Advancing Better Outcomes for All Children: Reporting Data Using a Racial Equity Lens](#)

Justice

Share data to reduce the burden of duplicate data collection.


Some people and communities are consistently the targets of data collection and study, sometimes from organizations seeking the same information for similar purposes. Siloed data place an additional—and potentially unnecessary—burden on community members. Sharing nonconfidential data, when it is unlikely that it could lead to harm or add risks, may reduce the burden that individuals and communities experience from data collection.

 **Resource:** [How to Appraise and Select Research Data for Curation](#)

Justice

Return data and research results to community members in a form they can use.

Analysts have the power to disrupt the dynamic of people's having no ownership of and deriving no utility from what their data have produced. Ensuring that the results are communicated in a way that community members can use and understand is an important step toward equity. In a research context, publishing in open access journals and providing action-oriented digests in line with community members' interests are two options. Other entities, including governments, may put data on open access portals or publish data in community-responsive ways for community members to use.

 **Resource:** [Why Am I Always Being Researched?](#)

DISPOSITION

Disposition is the stage at which people destroy or archive data, either completely or partially.

Justice

Empower individuals to order the destruction of their data.

Data stewards and managers do not “own” data more than the people whose lives are represented in them. To the extent possible and practicable, respecting people's wishes about the destruction of their data at any point during or after collection centers their concerns and welfare. Having a clear process for requesting the destruction of one's data and clear guidelines around honoring the requests are two ways to improve the formal process.

 **Resource:** [Best Practices for Data Destruction](#)

Respect for persons

Be transparent about what the plans are for the data after the project concludes.

Disclosures about disposition include information such as who has access to data after a project ends and whether, how, and when data will be destroyed. Many disclosures are written to protect organizations from legal liability, not to protect the interests of the people who provided data. Using plain language about data destruction is key so that people can decide whether they want to provide data at all, want to ask that their data be destroyed, or want only certain people to have access. This is particularly true if the data are to be shared with, sold to, or held by an indeterminate number of people whose purposes are unknown.

CONCLUSION

Data stewards, managers, and users in philanthropy, government, research, and beyond all have a duty to manage and use data responsibly, and we believe that responsibility includes incorporating equitable principles and practices throughout the data life cycle. What that looks like at different organizations and in different roles will vary. The examples in this document provide a starting point for imagining the possibilities. There is still much work to be done, and we invite you to build on these principles to figure out what works in your circumstances. Your work to include equity in data practice is essential to fostering new data norms that promote a more just data environment, which will help us all build strong communities for everyone.

ABOUT THE PROJECT

Elevate Data for Equity provides knowledge and tools for using data to advance equity and community health. Our publications and curated resources guide philanthropy, researchers, and local organizations as they build data capacity and critically examine their own data practices. The project draws on insights from the National Neighborhood Indicators Partnership and was funded by the Robert Wood Johnson Foundation. For more information, see the project website at <https://www.urban.org/elevate-data-for-equity>.

ACKNOWLEDGMENTS

This brief was funded by the Robert Wood Johnson Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission. In particular, we appreciate the encouragement and thought partnership of our program officers, George Hobor and Oktawia Wójcik, throughout the development of this brief and the careful review and insights from their colleague Maryam Khojasteh.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at [urban.org/fundingprinciples](https://www.urban.org/fundingprinciples).

We are thankful to our Urban Institute colleagues Peter Tatian, Marla McDaniel, Kathy Pettit, and Leah Hendeby who provided helpful perspectives and comments that greatly improved the brief. Meghan Ashford-Grooms of the Urban Institute improved the document with her editing, and Brittney Spinner elevated the brief with her design.

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Copyright © June 2020. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.