Education-to-Workforce Indicator Framework
Chapter V. Data equity principles

Framework excerpt

This file contains Chapter V of the Education-to-Workforce Indicator Framework. This chapter includes seven leading principles for centering equity throughout the data life cycle and supporting education-to-workforce systems to use data in service of greater equity. The full framework includes five chapters:

I.   Introduction and approach
II.   Indicators and metrics
III.   Disaggregates
IV.   Evidence-based practices
V.   Data equity principles

A. Overview

Working with data involves making decisions with equity implications. More than ever, government agencies, community organizations, and foundations use data to inform decisions about how best to promote more equitable education, workforce, and other policy outcomes for priority communities. However, organizations must be intentional in their use of data. How we collect, access, analyze, and report data can have serious and potentially detrimental impacts on individuals and communities, especially those already most marginalized, such as Black and Indigenous people, if we do not apply proper care and consideration. To counter these risks, data equity principles seek to ensure data are meaningful, accessible, and actionable for communities too often left out of data-driven decision making processes. This resource provides a synthesis of seven leading data equity principles that data users should apply throughout the data life cycle. It serves as a starting point, offering practical recommendations and additional resources for data users to approach education-to-workforce (E-W) data through an equitable lens and use data safely and securely.

> **Key terms**
>
> - **Asset framing:** Using language that focuses on the strengths, rather than deficits, of individuals or communities. Asset framing is the opposite of deficit framing.
>
> - **Community:** A place, institution, or group that includes individuals with similar characteristics, interests, or experiences (such as a neighborhood, school, or church).
>
> - **Data:** Distinct pieces of information, usually collected, stored, and processed in a way that is concordant with a specific purpose. They can be either quantitative or qualitative.
>
> - **Data users:** Individuals within organizations who collect and analyze data to inform decisions. These can include policymakers, administrators, educators, community leaders, and researchers, among others.
>
> - **Disparities:** Documented differences in outcomes between groups.
>
> - **Equity:** Just and fair inclusion into a society in which all can participate, prosper, and reach their full potential. Equity is achieved when structural barriers based on race, ethnicity, gender, sexual orientation, zip code, class, disability, and other factors are dismantled such that an individual's background and identities no longer predict their outcomes in life.
>
> - **Inequities:** The conditions that arise when policies, practices, attitudes, or cultural messages make it harder for some individuals—and easier for others—to fully participate, contribute, and take advantage of opportunities and resources based on their identities and background traits. Inequities are apparent when identities or background traits such as race, ethnicity, gender, sexual orientation, zip code, class, or disability statistically predict outcomes.
>
> - **Priority communities:** In the context of the E-W Indicator Framework, priority communities are identified as Black, Indigenous, and other communities of color and/or communities experiencing poverty. Priority communities may differ depending on the context and locale in which the framework is used.
>
> - **Proximate leaders:** Community advocates that share similar values and experiences of others within their community and are respected by community members as leaders and representatives.

B. What is data equity, and why does it matter?

Data can be a powerful tool when used purposefully and equitably. Data can empower practitioners, policymakers, and community members to make better, more informed decisions that are grounded in evidence, but they can also reinforce deficit narratives, biases, and other long-standing structural inequities when used inappropriately. To effectively assess and address disparities along the pre-K-to-workforce continuum, we must not only have access to more and better data, but also be deliberate in how we use those data. Whether intentionally or unintentionally, data can be misused and misinterpreted, sometimes causing harm to communities already most marginalized. Thus, we must be aware of these risks and apply an equity lens to every phase of the data life cycle.

Historically, E-W data have been used in both harmful and helpful ways, both to reinforce inequities or advance equity. Disaggregated education data have shined a light on the needs of particular groups of students, informing the passage of landmark policies such as the Elementary and Secondary Education Act, which established the Title I program to provide funding to schools with a high percentage of students from low-income households.[1510] At the same time, data on disparate academic outcomes, often referred to as "achievement gaps," have been used to argue the inferiority of specific racial groups, primarily Black and Indigenous people, and reinforce deficit-oriented beliefs that blame individuals rather than the systems that generate advantages for some groups and not others.

Today, algorithms built on E-W data are used in an array of applications that can positively or negatively affect individuals depending on their use. For instance, schools that have implemented Early Warning Intervention and Monitoring Systems to identify students at risk of not graduating for additional support have reduced chronic absence and course failure rates more so than schools without such data systems.[1511] But unintended consequences can also occur: after in-person exams were canceled due to the COVID-19 pandemic, the International Baccalaureate program's decision to use a data algorithm to predict students' grades resulted in systematically lower scores for high-achieving students from low-income households who had expected to earn college credit and save money on tuition.[1512]

These examples illustrate that data are not inherently neutral; like any tool, they require thoughtful use to achieve the intended goals. Using data in service of equity goals means that at every stage of the data life cycle, users must think critically about both the possible risks and possible benefits data might bring to the communities that provide data yet too often are left out the decision-making processes their data is ultimately used to inform. Data equity principles offer necessary guidelines for data users to ensure data are meaningful, accessible, and actionable for priority communities—thereby minimizing the risk of harm while maximizing the potential to promote greater equity through data.

C. Who is this resource for (and how should it be used)?

A growing number of resources offer guidance on how to work toward the goal of data equity, though it can be difficult to know where to begin. Many resources focus on a particular audience (such as researchers), phase of the data life cycle (such as visualization), or data application (such as results-based accountability). However, they share a set of underlying principles. This resource synthesizes seven leading data equity principles to apply throughout the data life cycle that are relevant to different types of data users and data projects.

Key audiences for this resource are education and workforce policymakers, administrators, educators, community leaders, and researchers who use data to diagnose disparities, implement evidence-based decisions, and evaluate the impact of policies, programs, and investments to address those disparities. It is a companion to the E-W Framework, commissioned by the Bill & Melinda Gates Foundation to encourage greater cross-sector collaboration and alignment across local, state, and national data systems by promoting the use of a common set of metrics and principles to assess and address inequities along the pre-K-to-workforce continuum. Through improved data systems *and* practices, organizations will be better poised to support the individuals least well served by education and workforce systems to achieve economic mobility and security.

This resource is intended to serve as a starting point. Applying data equity principles in practice can be complex, and best practices can take many forms depending on the specific context; this synthesis is only one resource data users should consult. After a summary of the seven core data equity principles, we provide overviews that further explain each principle and offer examples of how to apply it along the data life cycle, reflection questions and potential pitfalls for data users to consider, and additional resources to consult for more in-depth guidance. We encourage readers to refer to these original sources to dive deeper into the principles and associated best practices.

D. How was this resource developed?

This resource draws on data equity considerations gleaned from multiple sources, including leading publications by data equity experts and input from partners involved in E-W data systems. We began by conducting a literature review to gather information on how data equity principles are currently defined and applied in practice. Next, we presented an initial synthesis of this literature to a diverse range of partners, including education and workforce policymakers and data strategists, researchers, equity advocates, and parents and educators who make—and feel the effects of—data-driven decisions. This two-pronged approach incorporates scholarly, practitioner, and lived-experience perspectives into the data equity principles described in this resource.

Literature review

Using a targeted, iterative search strategy, we identified 32 publications that discuss guiding principles and best practices for centering equity in research or different phases of the data life cycle (see Appendix D for a complete list of sources). We summarized the common themes in these sources, which we then coded and synthesized down to seven core data equity principles that undergird the recommendations in the different source publications. Through the literature review, we also identified six key phases of the data life cycle during which data users should apply these core principles (Exhibit V.1).
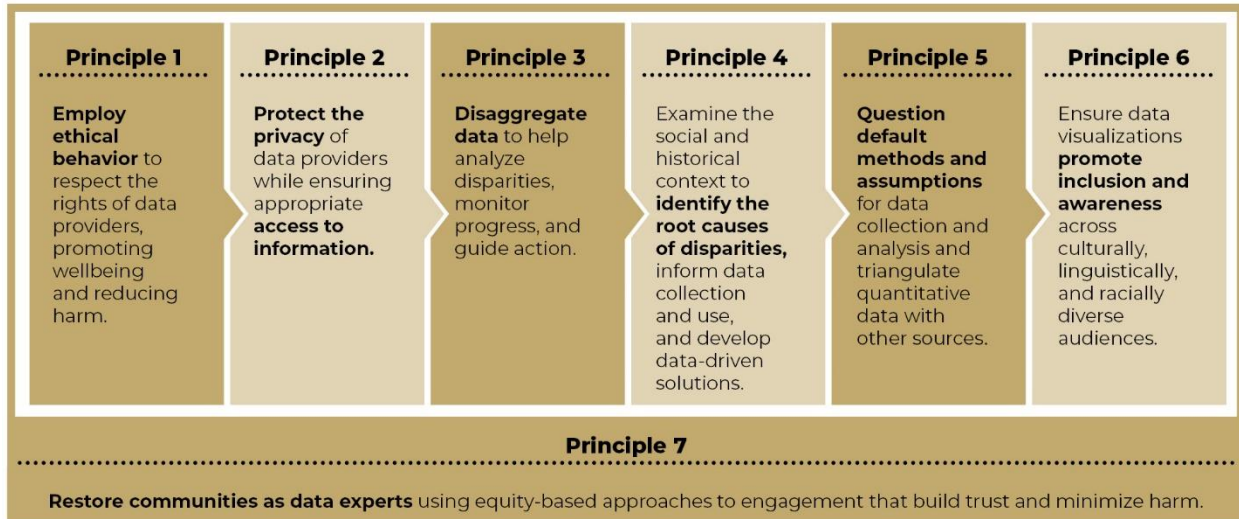
Exhibit V.1. The data life cycle

Contributory input

Following this initial analysis, we solicited feedback from a range of people connected to E-W research, advocacy, policy, and practice. This included the External Advisory Board of 18 E-W data experts and leaders and the internal working group of 10 Bill & Melinda Gates Foundation program officers, as well as members of five collective impact initiatives that lead advocacy efforts in their communities. (See the Introduction chapter for a complete list of individuals and organizations consulted.) During independent sessions with these groups, we solicited targeted feedback on the components of the E-W Framework, including this companion resource. Partners surfaced important gaps in current data systems and practices that too often omit contextual, system, and institutional factors that perpetuate inequities and leave out the communities most affected by the decision-making process. Their input informed the seven core data equity principles highlighted in this synthesis, as well as the guidance to implement them.

E. Seven core data equity principles

Below, we summarize the seven core principles for equitable data use (Exhibit V.2). The order in which they are listed is not indicative of their relative importance or priority—each principle must be put into action to achieve data equity. In particular, engaging community members as data experts (Principle 7) is critical to successfully implementing all of the other principles and meeting equity goals. Following this brief overview are three-page guides of each principle that include additional details, examples, and recommendations to guide their practical application throughout the data life cycle.

Exhibit V.2. Data equity principles

Principle 1	Principle 2	Principle 3	Principle 4	Principle 5	Principle 6
Employ ethical behavior to respect the rights of data providers, promoting wellbeing and reducing harm.	**Protect the privacy** of data providers while ensuring appropriate **access to information.**	**Disaggregate data** to help analyze disparities, monitor progress, and guide action.	Examine the social and historical context to **identify the root causes of disparities,** inform data collection and use, and develop data-driven solutions.	**Question default methods and assumptions** for data collection and analysis and triangulate quantitative data with other sources.	Ensure data visualizations **promote inclusion and awareness** across culturally, linguistically, and racially diverse audiences.

Principle 7

Restore communities as data experts using equity-based approaches to engagement that build trust and minimize harm.

PRINCIPLE 1: Employ ethical behavior to respect the rights of individuals who provide data, promote greater equity and well-being, and minimize the risk of harm.

Data users must evaluate data practices to determine whether they have the potential to contribute to greater equity, as opposed to reinforcing the status quo or even causing harm to communities already most marginalized. They must question whether they are addressing the underlying structural factors that perpetuate inequities, respecting the dignity and autonomy of all individuals, and maximize benefits while avoiding harm. At the outset of any data project, decision makers should identify and communicate the funding source and funders' priorities, the types of decisions the data project will inform, the data project's stated public benefit and equity goals, whether the data project meets the needs and addresses the concerns of the intended beneficiaries, and whether the data project could lead to unintended consequences or have racial equity implications. Decisions relying on data algorithms should be closely reviewed to ensure they do not have discriminatory or other unjust impacts. Involving community members in data governance, institutional review, and advisory structures can help achieve these goals.

PRINCIPLE 2: Protect the privacy of individuals who provide data while ensuring appropriate ownership and access to information.

Data users must seek the consent of individuals and recognize them as the owners of their data. Acknowledging that data represent the lived experiences of individuals, protecting data from improper use and exposure, and returning the data to community partners are all critical to promoting equity and earning public trust. Data users must follow data privacy laws and respect data sovereignty, for example, of Native American Tribes. Data users should consult the individuals providing data to determine who can securely obtain, view, or use data and for what purposes, weighing the risks and benefits of both restricting and opening access to data. Individuals should be allowed to access their personal data, correct data about themselves, and opt out from certain uses of their data. Decisions around data access can be made by a governance body that represents individuals who provide their data, including proximate leaders who authentically represent affected communities.

PRINCIPLE 3: Disaggregate data on both outcomes and system conditions to analyze disparities, monitor progress, and guide action.

Data users must acknowledge the diversity of experiences among priority communities to uncover disparities that can be hidden in aggregate data. Data analysis may require multiple levels of disaggregation to capture the intersectional nature of individuals' lived experiences. Thus, data users must collect data on multiple relevant background characteristics, guided by a contextual and theoretical understanding of root causes to avoid perpetuating existing stereotypes and deficit narratives. The E-W Framework offers guidance on key disaggregates to consider. In addition to disaggregating outcome data, data users should break out data on E-W and adjacent system conditions (such as funding) to reveal other underlying disparities.

PRINCIPLE 4: Examine social and historical contexts to identify root causes of disparities, inform data collection and use, and develop data-driven solutions.

To address disparities along the pre-K-to-workforce continuum, data users must understand the local social and historical context behind these disparities. Data users must examine data on structural conditions; learn about relevant past policies, programs, and institutions and how they may have

promoted or perpetuated racial inequity; and understand what members of priority communities see as the barriers to achieving equitable outcomes. Direct engagement with people with lived experience is key to conducting reflective root cause analyses focused on identifying systems drivers of disparities—not symptoms—and solutions to dissolve them.

PRINCIPLE 5: Question default methods and assumptions for data collection and analysis and triangulate quantitative data with other sources.

Data users must critically examine their methods and assumptions for collecting and analyzing data to ensure they do not inadvertently reinforce historical biases, deficit narratives, and power imbalances. Quantitative methods are sometimes viewed as being inherently objective, but data users must be attentive to these risks and question their own motives and biases, where the data came from and what they might leave out, and who they see as the experts on the data. When seeking to answer questions, data users should consider triangulating quantitative methods with other approaches to inquiry, such as collecting qualitative data from interviews or focus groups to capture additional insights or designing community participatory action projects that privilege community voice and participation. Gathering multiple sources and types of information can help counter the bias in any one data source.

PRINCIPLE 6: Ensure data visualizations promote inclusion and awareness across culturally, linguistically, and racially diverse audiences.

Data users must approach visualization with thoughtful consideration to the lived experiences the data communicate and to every detail used to present that information—including labels, colors, ordering, graphics, and icons—to ensure it is accessible to multiple audiences and does not reinforce stereotypes and deficit narratives. Information on the source of the data, when and why they were collected, and who they represent should accompany visualizations. This and other contextual information (for instance, centering the structural causes behind disparate outcomes being shown, either though narrative text or additional data on system conditions) can be key to ensuring that readers do not misinterpret or misuse data visualizations.

PRINCIPLE 7: Restore communities as data experts using culturally responsive approaches to engagement and co-creation that support equitable data use.

Community partners are a vital resource for data users. As illustrated in all of the principles, engaging community members with lived experience is key to centering equity throughout the data life cycle. Data users should follow best practices for effective community engagement, which include defining clear expectations and roles at the outset of a data project; recognizing and examining the power imbalances between decision makers and community members; building in enough time for community members to engage meaningfully in the project; allocating resources to equitably compensate community members; and avoiding the risk of exploiting, tokenizing, or retraumatizing them. As much as possible, data projects should build community capacity to use data to advocate for change, for example, by co-designing projects that reflect the community's values, histories, culture, perspectives, and voice.

PRINCIPLE 1: Employ ethical behavior to respect the rights of individuals who provide data, promote greater equity and well-being, and minimize the risk of harm.

Ethical behavior requires data users to evaluate data practices to determine whether they have the potential to contribute to greater equity, as opposed to reinforcing the status quo or even causing harm to communities already most marginalized, such as Black and Indigenous people. It requires data users to consistently challenge ideas, practices, or policies that fuel systemic racism. To combat systemic racism means to challenge the notion that differences between racial groups are simply inherent, rather than understanding that racial disparities are a product of longstanding oppressive systems and policies. Data users must question whether they are addressing the underlying structural factors that perpetuate inequity, respecting the dignity and autonomy of all individuals, and maximizing benefits while minimizing the risk of harm.

Although Institutional Review Boards (IRBs) determine whether ethics are upheld in research,[xxxi] in practice IRBs are not well equipped to perform deep reviews that center the concerns of marginalized groups to advance racial equity. For example, IRBs have allowed people of color to be systematically underrepresented in clinical trials, even when they are most affected by the health conditions being studied.[1513] In addition, many data projects occur in settings with little or no ethical oversight. Data users must carefully assess data projects' potential risks and benefits to the well-being of individuals and society at large to avoid being extractive and exploitative. Data users must weigh the risks and benefits holistically, with an eye toward the groups that might be differentially affected to ensure both risks and benefits are distributed fairly, and racial equity is being promoted.

Data users should be attentive to uses of data that carry a high risk of causing harm, such as algorithms, or data-based decision tools, that may lead to discriminatory practices. Algorithms reflect the biases of the people who develop them and of the underlying data. If considering using an algorithm to inform decision making, data users must ensure transparency, assess algorithmic bias, and determine the potential positive and negative consequences of applying the algorithm in practice. Decisions based on a data algorithm should always be reviewed by humans, and affected individuals should have the ability to contest the decision. Data users should also be attentive to minimizing the amount of data collected on sensitive topics (for example, mental health) and rigorously protecting personally identifiable information.

At the outset of any data project, decision makers should identify and communicate who is funding the project and what their priorities are, the types of decisions the data will inform, the data project's stated public benefit and equity goals, whether the data project meets the needs and addresses the concerns of the intended beneficiaries, and whether the data project could lead to unintended consequences or have racial equity implications (good or bad). They must engage the groups of people whom the data project might affect to make these determinations, be responsive to their feedback, and ensure transparency.

Community engagement is especially critical if the project could have serious or disproportionate impact on marginalized groups or those facing multiple barriers. Involving multiple partners, including proximate leaders from affected communities, in data governance, institutional review, and advisory structures, can help data users ensure the project is successful in promoting equity and well-being.

[xxxi] Ethical principles of research are described in the Belmont Report, which guides human subjects' protections in research (but does not have a racial equity lens).

Ideally, community members can co-create project goals and plans with proximate leaders to ensure the data are meaningful and actionable to them and counter existing power structures. These bodies should be convened early and offer continued input and oversight throughout the data life cycle.

> **The importance of transparency in ethical data use**
>
> Mount Saint Mary's University, a small, private college in Maryland, made the news in 2016 after a plan to use student data to boost retention rates became public. New students would have to take a survey that the school would use to predict their likelihood of dropping out; students with a high probability would then be encouraged to unenroll before they were counted in the retention data that colleges report to the federal government. Mount Saint Mary's did not disclose to students that their survey responses could be used to encourage them to leave (Ekowo & Palmer, 2016)—a major ethical breach. In contrast, other colleges, such as Georgia State University and Temple University, have successfully used predictive analytics to improve graduation rates by involving students and staff in the process. Transparency is at the heart of using data ethically and equitably, allowing for greater oversight and accountability.

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Context-setting	Hold listening sessions with community members to learn what types of data projects the community thinks are relevant to improve their lives. Consider the impacts of structural racism on the priority community, and listen to the stories of community members to identify ways the work could be beneficial to them. Examine the results of past data projects, including past approaches to centering equity, to identify strengths and areas for improvement.
Planning	Establish a governance or review body with representation from multiple contributing groups, including proximate leaders from affected communities. Convene this body to agree on the goals of the project, identify risks and benefits, develop mitigation strategies, and inform decisions at each phase of the data cycle. Consider formalizing a commitment to ethical data use by drafting a social impact statement[1514] that outlines how to put principles into practice.
Collection	Minimize the collection of sensitive and personally identifiable information unless it is critical to achieving the project's intended benefits. Eliminate the collection of any nonessential data to minimize burden on individuals. Individuals, especially those in marginalized communities, may perceive the collection of unnecessary personal information as over-surveillance and question whether the data collection has hidden purposes.
Access	As appropriate, securely share data with partners to reduce the burden of duplicate data collection (see Principle 2 for additional considerations on data privacy and access). Communicate policies on data storage, access, and use in lay terms.
Analysis	Clearly describe the methods and algorithms used to analyze the data, their potential for inaccuracy and bias, and how they will be used to inform decision making. Seek out and incorporate communities' interpretation of the data.

Key phases for this principle	Example applications
Reporting	Return data and research results to community members in a form they can use. Create channels to report grievances. Publicly disseminate the results of the analysis and invite others to build on the research in an ethical manner that will produce continuous benefits to the community. Accurately identify the strengths and weaknesses of the data.

Reflection questions

- Who would benefit from or be burdened by the data project? Are both benefits and burdens shared equitably?

- What are the potential risks of the project versus the risks of not proceeding with it?

- Could you modify the project to enhance positive impacts or reduce negative impacts?

- Are governance and oversight mechanisms in place? Do they include community representation?

- How will you know whether the intended benefits to the community were achieved?

Be on the lookout ⚠️

"Early warning" and other predictive indicators can be powerful tools to help E-W systems support students earlier and more effectively. However, they should not be used for increased monitoring or punitive action. Data users must be aware that biases in the inputs used to form predictions can perpetuate stereotypes and even lead to discriminatory treatment. For example, although past suspensions are predictive of high school graduation, they also reflect racial bias in school-based disciplinary actions.[1515,1516] Thus, algorithms should never override the judgment of individuals. Balancing information from the algorithm with the judgment of practitioners, students and parents, and other qualitative or contextual data can help ensure equitable outcomes are achieved.

Additional resources

- [Principles for Advancing Equitable Data Practice](). This brief by the Urban Institute describes the Belmont Report's ethical principles and offers examples of practices and resources to integrate the principles throughout the data life cycle with an equity lens.

- [The Data Equity Framework](). This framework from We All Count identifies key equity-impacting decision points in data projects and offers practical tools for developing and implementing ethical data projects that center equity.

- [A Toolkit for Centering Racial Equity Throughout Data Integration](). This toolkit by Actionable Intelligence for Social Policy includes chapters on "Racial Equity in Planning" and "Racial Equity in Algorithms/Statistical Tools" which describe positive and problematic practices with ethical implications, as well as citing brief case studies.

- [Forum Guide to Data Ethics](). This report by the National Forum on Education Statistics offers nine "canons" of data ethics in education, along with real-life examples and resources to implement these canons.

- Racial Equity Considerations and the Institutional Review Board. This Child Trends blog post describes why racial equity matters in IRB submissions and offers suggestions for applying an anti-racist lens when submitting to an IRB.

PRINCIPLE 2: Protect the privacy of individuals who provide data while ensuring appropriate ownership and access to information.

Data privacy policies protect the right of individuals to maintain control over their data. They include a combination of federal, state, and local laws—including the Family Educational Rights and Privacy Act (FERPA)[1517]—and institutional policies. Most policies focus on protecting personal information—or information that is important to an individual (even if it does not personally identify them)—and regulating data access and use, thereby limiting emotional, financial, and even physical harm that can result from data privacy breaches. Although privacy considerations are critical, it is also important to understand and honor data ownership. Data users must acknowledge that data providers are data owners that consent to the use of their data.

Data privacy policies have evolved in recent years to better reflect that data systems do not "own" data more than the people whose lives are represented in them. In 2018, the European Union passed the General Data Protection Regulation,[1518] which gives European residents the right to know, access, update, erase, and restrict the types of data collected on them. Since 2020, the California Consumer Privacy Act (CCPA)[1519] requires businesses (including for-profit education service providers and for-profit universities) to obtain parent or guardian consent before collecting data from California's children and to delete data upon request, among other things (CCPA has inspired similar laws in other states). A common feature of these laws is that they grant individuals the ability to update, delete, or opt out of all or specific applications of their data at any point during or after collection. Even if not mandated by law, E-W data systems should have a clear process for accepting these requests and clear guidelines around honoring them.

Data users should consult community members to determine data access guidelines and practices, weighing the risks and benefits of both restricting and opening access to data. Data access refers to who can securely obtain, view, or use data, and for what purposes. There are legal, practical, and equity considerations for determining data access, which can range across contexts. For example, sharing administrative data with E-W system partners or researchers can increase the risk of a data breach, yet *not* sharing data can make it more difficult to understand and address a problem of practice, at least without duplicating data collection efforts that burden communities. At a minimum, communities should have access to their own data (abiding with any privacy or confidentiality rules). But access is different from ownership. To shift power dynamics and honor communities' own goals and visions, communities should have the right to govern the collection, ownership, and use of their data. This is a key principle of Indigenous data sovereignty, for example.[xxxii]

E-W data systems should establish a participatory governance structure that includes representation from the affected communities to determine which data are open, restricted, or unavailable and—as with requests from individuals about their own data—develop a clear process for accepting and approving requests from potential data users. After a project ends, data users should consider secure methods by which they can return data (for example, in aggregate form) to the communities, the data owners, to allow continued or future use of their data for other purposes.

[xxxii] See this 2018 resolution from the National Congress of American Indians: "Support for U.S. Indigenous Data Sovereignty and Inclusion of Tribes in the Development of Tribal Data Governance Principles."

The real risks of data breaches

The Government Accountability Office (GAO) discovered 99 data breaches in 281 school districts from July 2016 to May 2020. The breaches affected thousands of students and parents, exposing sensitive data such as special education records, test scores, phone numbers, and Social Security numbers. School staff, students, cybercriminals, and vendors were all responsible for various data breaches, which were both intentional and accidental. Citing the risks to students' physical, emotional, and financial well-being, the GAO recommended that schools review and follow data privacy laws, provide data security trainings, require vendors to configure data systems adhering to the Federal Trade Commission's "Start with Security Guide," or take an annual Nationwide Cybersecurity Review self-assessment.

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Context-setting	Review federal, state, local, or Tribal data privacy laws and policies that apply. Determine whether you need memoranda of understanding, data-sharing agreements, or consent to collect or share data.
Planning	Develop a list of data elements to collect and any linked data sets, as well as how you will store data, who will have access to data, how you will use data and for how long, and what you will do with the data after analysis is complete. Establish a governance body with representation from multiple contributing groups, including proximate leaders from affected communities. Convene this body to develop clear processes and guidelines for accepting and approving requests from individuals who provided their data and potential data users.
Collection	Communicate data privacy and security processes when collecting data. Seek informed consent even if not required. Only collect data that are necessary and have been approved.
Access	Store data in a secure location that is only accessible to authorized users. Ensure storage systems have the proper protections (such as locks, encryption, and passwords). If you share data, ensure they are transmitted through secure methods. Train those with access to data on relevant laws and best practices. Practice data minimization; only give users access to the minimally necessary data elements and data sets. Ensure individuals who provide data can access, update, and delete their data upon request. Upon project completion, discard or return data as directed or previously established by individuals who provided the data.
Reporting	Maintain confidentiality of participants in reporting. Do not name individuals without permission, share a combination of data points that could lead to an individual being identified, or report data on very small sample sizes that could risk identification. Delete data when no longer in use for the intended purposes.

Reflection questions

- Beyond federal data privacy laws such as FERPA, which state, local, or Tribal data privacy laws or policies apply to you?

- What procedures have you established to enable individuals to access, update, or delete their data, if requested?

- If many people opt out of data collection, why have they done so? How can you use their feedback to inform and redesign data collection efforts to minimize conflict and harm?

- What will you do with the data after analysis and reporting? Can you share the data back with communities? How can the individuals who provided their data inform your decision?

Be on the lookout ⚠️

Data sharing between organizations can give users access to additional data elements needed to assess and address disparities and reduce the data collection burden on individuals; however, it comes with its own risks. Any time data are shared, users must follow data governance policies by establishing a memorandum of understanding or data-sharing agreement and reviewing any consent documentation to ensure data sharing is permissible. Both parties must transmit the data securely and clearly track the data lineage—where the data came from and where they're going. Never share data with third parties (whether businesses, researchers, law enforcement, or other government agencies) or use for other purposes without permission.

Additional resources

- [Roadmap to Safeguarding Student Data](#). This Data Quality Campaign implementation road map for state education agencies overviews relevant data privacy laws and best practices for transparency, governance, and data protection procedures.

- [A Path to Social Licence: Guidelines for Trusted Data Use](#). Data Futures Partnership offers eight guidelines for data use related to data value, protection, and choice. Although some of the guidelines are specific to New Zealand and its Tribal communities, many are universally applicable.

- [A Toolkit for Centering Racial Equity Through Data Integration](#). The chapters on "Racial Equity in Data Collection" and "Racial Equity in Data Access" by Actionable Intelligence for Social Policy address positive and problematic policies related to data privacy, as well as cite brief case studies.

- [Indigenous Data Governance: Strategies from United States Native Nations](#). This journal article by Russo Carroll et al. explains the concepts of Indigenous data sovereignty and governance, and describes the value and challenges of shifting authority over Indigenous data to Indigenous peoples. The article includes Tribal case studies and discusses relevant federal laws and Tribal organizations.

- [Envisioning a New Future: Building Trust for Data Use.](#) This resource, developed by the Urban Institute for the Data Funders Collaborative, describes approaches to building trust for collection and use of data, such as ways to expand and control data access and improve systems for consent and transparency. It includes a list of additional resources for data use and integration.

PRINCIPLE 3: Disaggregate data on both outcomes and system conditions to analyze disparities, monitor progress, and guide action.

Data users must acknowledge the diversity of experiences among priority communities to uncover disparities that can be hidden in aggregate data. Data analysis often starts by measuring outcomes for broad populations of individuals, but results can vary—sometimes significantly—across certain populations or groups with unique experiences and histories. Taking a passive stance in data analysis can lead data users to draw different conclusions. Without disaggregation, they may miss the opportunity to identify, address, and monitor disparities. The E-W Framework offers guidance on 25 key disaggregates data systems should collect, including race and ethnicity, gender, income level, disability status, English proficiency, and lesbian, gay, bisexual, transgender and queer or questioning (LGBTQ) status.

Though data systems must collect or link data on multiple relevant background characteristics, *which* factors are analyzed through disaggregation and *how* they are analyzed depend on the local context. Data analysis may require more than one level of disaggregation to capture the intersectional nature of individuals' lived experiences. For example, a school district might explore whether high school graduation rates differ for students with disabilities by race. In contexts with smaller populations, disaggregating across multiple levels is not always feasible as subgroup sizes grow smaller with each level of disaggregation, making it harder to reliably compare trends over time. However, data users must still consider the experiences of smaller groups, such as American Indians and Alaska Natives, and not simply default to grouping them under an "other" category that does not receive careful attention.

Decisions about how to disaggregate data should also be guided by a theoretical understanding of a problem of practice and potential root causes to avoid perpetuating existing stereotypes and deficit narratives or framing that advertently or inadvertently blame particular groups rather than systems for disparate outcomes. In addition to disaggregating outcome data, data users should break out data on E-W and adjacent system conditions to reveal other underlying disparities. For example, system conditions such as access to school support staff may be relevant to the graduation rates of students with disabilities, and these indicators should also be disaggregated further by race. However, disaggregation alone is not enough to reveal causes or solutions for inequities, as described in Principle 4 on examining social and historical contexts to identify root causes of disparities and data-driven solutions.

> ## When "standard" disaggregation is insufficient
>
> Data users should consider whether standard categories commonly used to disaggregate data, such as broad racial categories, may not be appropriate for all groups and contexts. For example, an analysis of census data on four-year postsecondary degree completion by race would show that more than half of Asian Americans have a bachelor's degree or higher, the highest rate among any racial group. However, this rate masks significant variation within different communities of Asian Americans: for instance, less than 15 percent of Laotian Americans obtain bachelor's degrees. Disaggregating data by both race and detailed ethnicity categories shows that certain groups of Asian Americans, including Laotian, Cambodian, Hmong, and Vietnamese Americans, experience educational attainment on par with other minoritized groups. To put these differences into context, users should also collect and disaggregate data on potential root causes that drive educational attainment for different ethnic groups, such as their reasons for immigration, generational status, neighborhood resources, or access to financial aid.

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Planning	Work with community members to determine which characteristics to measure during data collection or to link into the data (if already available), and how to label these characteristics in data collection tools as well as eventual reporting (for example, Hispanic, Latino/a, Latinx).
Analysis	Disaggregate both outcome and systems data at multiple levels to illuminate any disparities. Include qualitative research or input from the community so that readers can contextualize disaggregated data with individuals' lived experiences and the root causes of any observed disparities.
Reporting	When reporting disparities by subgroup, connect these to the system and root causes, not people. Use data visualization to clearly communicate disparities while avoiding perpetuating deficit narratives (see Principle 6).

Reflection questions

- Who is or is not included within the categories representing the population of study?

- How can disaggregated data help us think about intersectional issues (for example, how outcomes might differ for Black boys versus Black girls)?

- Have we analyzed both outcome and structural disparities between subgroups and avoided placing blame or perpetuating stereotypes?

- When is it appropriate to compare data within versus between groups (for example, comparing outcomes for Latino high school graduates and Latino college graduates versus comparing outcomes for Latino and non-Latino college graduates)? Which comparisons would best answer your research questions and inform future action?

Be on the lookout ⚠️

Data users should tailor plans for disaggregation to each community and not simply report on mandated categories. For instance, defaulting to disaggregating data by just race and income would not provide much additional insight in a community comprised almost exclusively of Latino families

with low incomes. Depending on the community's local context and the problem of practice being considered, further disaggregation by factors such as English proficiency and newcomer status may reveal hidden disparities that systems should understand and address.

Additional resources

- [Disaggregated Data: Not Just a Box Checking Exercise](). This three-page brief by the Data Quality Campaign, Learning Heroes, and National Parent Teacher Association details what data disaggregation is, why it matters in K–12 education, which subgroups are required for disaggregation under the Every Student Succeeds Act, and how to communicate the value of disaggregated data to interested groups (including examples from multiple states).

- [The Essentials of Disaggregated Data for Advancing Racial Equity](). This Race Matters Institute blog post offers guidance on how far to go in data disaggregation, deciding which data to disaggregate, and presenting disaggregated data.

- [By the Numbers: A Race for Results Case Study](). This Annie E. Casey Foundation report shares two cases studies of how data users have disaggregated data to inform policies, practices, and decision making for their populations of focus.

- [The Importance of Disaggregating Data](). This short report by Safe Schools Healthy Students addresses the importance of disaggregating data (including examples), common disaggregates, and limitations of data disaggregation.

PRINCIPLE 4: Examine social and historical contexts to identify root causes of disparities, inform data collection and use, and develop data-driven solutions.

To assess and address disparities along the pre-K-to-workforce continuum, data users must understand the local historical and social context behind these disparities. Root cause analysis equips decision makers with the essential contextual knowledge needed to understand how disparities are produced, not only that they exist. Too often, data users analyze data on outcomes without deeply interrogating the structural causes of the disparities they observe, such as historical events, racist and other unjust policies, misinformed interventions, and oppressive social conditions. Without an understanding of these root causes, data projects and intervention strategies can fall short of creating lasting change and may even perpetuate racist structures.

Root cause analysis is a data-driven inquiry process with three overarching steps: identify a problem, identify root causes of the problem, and identify strategies to address the root causes. Data users must spend time developing an understanding of system conditions and other contextual factors that might be contributing to disparate outcomes, pulling data and information from existing sources, if available, to avoid duplicating efforts and placing undue burden on community members. Grounding data work in historical and societal context can also involve conducting an organizational reflection, equity audit, or environmental scan. An equity audit is a study of the fairness of an institution's policies, programs, and practices.[1520] Equity audit tools can help data users critically examine policies, programs, and practices that directly or indirectly affect students or staff related to their identity. An environmental scan involves gathering information about a community and its relationships to understand the systems and institutions in place that affect how people behave, and the landscape in which the community operates.

Direct engagement with people with lived experience is key to conducting reflective root cause analyses that seek to identify systems drivers of disparities—not symptoms—and solutions to dissolve them. After an initial assessment of disparities, data users should convene groups of people with different perspectives on the problem—such as practitioners, students, and parents from priority communities—to brainstorm possible explanations that, if addressed, ought to reduce or prevent disparities in the future. Groups should prioritize potential root causes until they reach consensus on a few of the most actionable factors most likely to drive disparities. This process should not only inform the development of solutions, but also decisions about which data to collect and analyze to further validate the hypothesized root causes and monitor progress.

> **Involving community to identify and address root causes**
>
> Disaggregated test score data for Marguerite Montgomery Elementary School in Yolo County, California, showed that students in the school's English-only program scored significantly lower than their peers in the two-way bilingual immersion program in every grade, regardless of whether students were emerging multilingual learners. The school held multiple staff and parent engagement activities in both Spanish and English to uncover the root causes of this disparity. They found systemic disproportionalities in the students enrolled in the two programs. They also learned that the school community valued bilingualism, and that research showed that students in dual language programs did as well or better than their peers in English-only programs. As a result, the school decided to transition into a fully dual immersion model, holding planning sessions that continued to engage both staff and community members as part of a new continuous improvement cycle ([California Department of Education, 2021](#)).

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Context-setting	Identify key historical events, policies, and processes that provide context for the observed present-day disparities. You can conduct an historical analysis through an equity audit, an environmental scan, or organizational reflection, such as a visual timeline activity that maps trends in outcome data against policies and other changes over time.
Planning	Vet research questions and data collection plans for a root cause analysis with the groups of people most affected by the identified problem of practice. Community members can provide input on whether the right problem of practice has been prioritized and which data points should be collected and from whom to explore its root causes.
Analysis	Engage multiple colleagues in dissecting the chosen problem by asking them to answer the question, "Why is this the case?" five times.[1521] Tools like a fishbone diagram[1522] or root cause tree[1523] can aid in this step. Focus on systems and structures, eliminating explanations that are not within the control of E-W decision makers, are not consistent with the available data, or cannot be tested. Reach consensus on the most likely and actionable root causes.
Reporting	Seek community reactions to and interpretation of findings to illuminate root causes not otherwise surfaced. Co-create action items—including potential data-driven solutions to address the root causes—to promote change through advocacy.

Reflection questions

- Who is affected—positively or negatively—by the disparity in question? Why? How?

- Do our analyses identify historical structures, policies or practices, and institutions involved? What social conditions contribute to the problem?

- Do our analyses go far enough, or are we attributing an equity disparity to contributing factors rather than root causes? Are there alternative explanations that fit better?

- What opportunities have we provided for community members to lead and drive contextual understandings to support project goals?

Be on the lookout ⚠️

Be careful not to mistake *contributing factors* for root causes. Contributing factors are conditions that allow the identified disparity to occur or persist. A root cause is a factor that prevents it from occurring if taken away. Removing a contributing factor (for example, expanding Advanced Placement course offerings) can improve disparate outcomes, but will not eliminate them. Addressing root causes (for example, educator bias, misplacement of Black students in noncollege preparatory courses) makes it more likely that solutions will be successful in promoting equitable change.[1524]

Additional resources

- [How to Embed a Racial and Ethnic Equity Perspective in Research.](#) This guide by Andrews et al. offers practical guidance to researchers and data users alike on how to dissect and use data through an equity lens. The authors pay particular attention to understanding the contextual and societal factors behind the issues of access and opportunity a community may face.

- Race Equity and Inclusion Action Guide. This Annie E. Casey Foundation resource provides guidance on key steps to advance and embed racial equity and inclusion in organizations. It provides questions to guide data users through a systems analysis of root causes of inequities and to identify strategies to address root causes.

- The State and District Role in Root Cause Analysis. This resource provided by the Office of Elementary & Secondary Education links to tool kits that state and district education agencies use to conduct root cause analyses while supporting school improvement efforts. It also offers guiding questions and facilitation tips for districts and states.

- How We Should Talk About Racial Disparities. This article by Spievack and Okeke discusses *why* and *how* researchers and data users can examine contextual factors to avoid perpetuating racist structures and eliminate bias in reporting.

PRINCIPLE 5: Question default methods and assumptions for data collection and analysis and triangulate quantitative data with other sources.

Data users must critically examine their methods and assumptions when collecting and analyzing data to ensure they do not inadvertently reinforce historical biases, deficit narratives, and power imbalances. Modern data collection and research methods are rooted in legacies of racial power imbalances and exploitative practices. Some lasting effects of these legacies include maintaining whiteness as the standard to which other groups are compared (for example, reporting Black-White and Asian-White gaps in outcomes) and over-relying on quantitative data, which can perpetuate stereotypes, without considering qualitative, contextual factors. Data teams that lack racial and ethnic diversity and varied life experiences, including experiences close to the community at the center of data projects, may reflect inherent biases. The makeup of data teams can lead to potentially misleading research questions, uneven power sharing, and assumptions of what data are "meaningful." By triangulating quantitative data with qualitative information and reexamining personal and institutional biases, data users can mitigate these risks.

Quantitative data alone are insufficient to illuminate the full picture of a community's experiences. Though often seen as objective, quantitative data can reflect the biases of the researchers and administrators who design data collection instruments and of the individuals who report the data (such as teachers and police). Relying solely on quantitative data can also remove pertinent institutional factors from analysis that reveal critical information. Using qualitative methods in addition to quantitative methods can more adequately capture why and how disparities exist, including root causes. Qualitative data sources include focus groups, interviews, observations, or long-form surveys. In some projects, it can be appropriate to employ community-based participatory research methods (CBPR)—one model that challenges traditional research structures. CBPR prioritizes collaboration between data users and community through equal partnership. Whatever methods data users choose, they must ensure data collection instruments are clear, unbiased, and speak to the experiences of community members by piloting questions and revising them accordingly.

The racial, socioeconomic, and cultural identities of data users implicitly influence the research questions they seek to answer, the way in which they collect data, and the methods through which they analyze and report them. Before a project begins, data teams should consider their team dynamics and characteristics and examine their individual and group implicit biases, for example, by using tools like the Implicit Association Test[1525] or an intentional reflection of how the team's experiences and motivations might differ from those of the priority population. In doing so, team members with less dominant identities should be able to opt out of potentially harmful spaces. Uncovering, acknowledging, and addressing personal and institutional biases at the outset can guide the team's approach to each phase of the data life cycle. For example, if a project involves employment data, the team can assess whether bias exists in its definition of "valid'" employment and adjust data collection or analysis plans to make the inquiry more inclusive. Exhibiting cultural competency and including a diverse team of data users with proximate experiences to the priority community increases the accuracy and ultimate benefit of the data work.

> **Child Trends initiative with PBS Kids**
>
> A 2019 Child Trends initiative with PBS Kids sought to develop family engagement programs in four communities. To ensure program designs were rooted in community needs, Child Trends launched a community assessment study as a first step. The team held an open discussion to consider how its experiences differed from those of the communities it planned to interview, including how bias might influence proposed interview questions. The team then repositioned interview questions to lead with the existing strengths in family engagement efforts, rather than gaps or weaknesses. Next, to challenge the norm of centering White, middle-class experiences and values as the standard for family engagement, the team employed a "360-approach" to understand the priorities in schools across the four communities. This approach involved interviews with educators, parents, and leaders of family groups. The strategy ensured the team did not just default to an approach that would not be useful to each community.

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Context-setting	At the outset of a data project, conduct an implicit bias test or group reflection activity among the proposed team to identify individual and institutional biases and discuss ways to mitigate them throughout the project life span. To increase cultural competency, learn about the history, power structures, and systematic barriers that exist in priority communities, as well as the community's prior experiences with data collection efforts. Continue questioning biases and assumptions in each subsequent phase.
Planning	Ensure data teams reflect diverse lived experiences, and in particular the experiences at the center of the data project. Consider which type of data collection or research model the project is proposing—traditional, community-engaged, or full community partnership. Examine whether the proposed approach and metrics inject any assumptions about the partner community, or whether they place undue burden on them. Pilot all data collection instruments, both qualitative and quantitative, with community members to ensure the instruments are culturally aligned to capture accurate and reliable data.
Collection	Employ qualitative methods, such as interviews, focus groups, town halls, narratives, or long-form surveys, to triangulate quantitative methods. Gathering data through a wide variety of sources strengthens analysis and can validate, contextualize, or challenge quantitative findings.
Analysis	Carefully consider whether findings perpetuate or reinstate a negative stereotype or deficit narrative. If findings meaningfully neglect institutional or systemwide factors, consider how community input might supplement the evidence to give a fuller picture.

Reflection questions

- What assumptions are built into the proposed data collection or analysis approach?

- Is the data team reflective of and close to the community whose data are being collected? If not, has the team conducted an implicit bias exercise or group reflection?

- Have efforts to examine the disparity in question existed in the past? Can you pull from those efforts and supplement quantitative data through qualitative exploration?

- Have you piloted research instruments or data collection prompts with members of the priority community? Do the instruments reflect assumptions about the priority community? Can they be repurposed using asset-based framing?

Be on the lookout ⚠️

Publicly available quantitative data sets often report measures of compliance, such as arrest and suspension rates. These "simple" measures may be cheaper and easier to collect, but can perpetuate stereotypes and deficit narratives if not analyzed with care.[1526] Data users should think closely about the metrics they choose and consider whether they are defaulting to using data that happen to be available, even if the resulting metrics are not as meaningful for the project's goals. When possible, data users should gather input from community partners when selecting data for collection and define metrics using asset-based framing. If the project must use a "simple" measure that relies on available data, data users should supplement it with other data points, including qualitative data, to help in interpretation.

Additional resources

- [The Equitable Evaluation Framework](#). The Equitable Evaluation Initiative's site offers a framework of principles to align evaluation practices with an equity approach, along with a suite of resources, reflection tools, and examples to help data users apply these principles.

- [Why Am I Always Being Researched?](#). This Chicago Beyond guide offers ways to authentically partner with and engage community members in selecting approaches and methods to data collection and analysis. The section "For Researchers" (p. 62) discusses specific probes to challenge internal and institutional biases in default methods.

- [Making Racial Equity Real in Research](#). This resource from the Greenlining Institute outlines promising and problematic practices throughout the data life cycle. The sections "Methodologies, Data Collection and Analysis Can Perpetuate Inequities" (p. 14) and "Lack of Cultural Competency of Researchers" (p. 15) caution against pitfalls and offer promising practices when launching data collection initiatives.

- [How to Embed a Racial and Ethnic Equity Perspective in Research](#). This Child Trends resource introduces a model for data collection through the lens of five equity principles, including that "researchers should examine their own backgrounds and biases." In addition, it offers guidance on qualitative and quantitative data collection and analysis.

- [Community Based Participatory Research](#). Chapter 36 of this University of Kansas guide on evaluation outlines principles and practice guidance for engaging in CBPR, an alternative to traditional research.

PRINCIPLE 6: Ensure data visualizations promote inclusion and awareness across culturally, linguistically, and racially diverse audiences.

An equitable approach to data visualization ensures data do not reinforce stereotypes and deficit narratives and are accessible to multiple audiences. Data visualization refers to the graphs, icons, pictures, colors, order, and labels used to represent patterns in data. Using visual representations to portray findings has the power to distill large amounts of evidence into digestible, visual narratives. However, if done without an equitable lens, visualizations can "otherize" particular groups, reinstate bias, and obscure findings for audiences without research backgrounds. Statistics are grounded in real people and communities. Data users have the power to reflect dignity, empathy, and respect for those narratives through equitable visualization practices.

Equitable data visualization employs colors, labels, ordering, graphics, and icons in consideration of the lived experiences that data communicate to the intended audience. In addition to following federal accessibility guidelines,[1527] data users should carefully consider how visualization elements might reinforce stereotypes. For example, graduated color palettes imply a scale, so should not be used for categorical data, such as listing racial groups. Similarly, choosing a male-presenting icon to depict a school principal can reinforce a stereotype that female-presenting individuals are not suited for leadership roles. Titles and labels should use person-first language, such as "people with disabilities" instead of "disabled people." Asset-based framing can also shape how readers view statistics and the people behind them—for example, by showing the number of students "meeting benchmarks" as opposed to the number of students "below grade level." As another example, data visualizations should not default to using White students or individuals as the benchmark for other groups, but must be mindful of which comparisons are most clear and meaningful.

Equitable data visualizations must keep their audience in mind, which should include the greater community from which the data were gathered. Using overly technical and jargon-filled visualizations is not only dismissive of some audiences, but also removes data ownership from communities and puts power back in the hands of researchers and decision makers. Accessibility, however, does not imply oversimplification. Data users must ensure the reader has the context, references, and annotations needed to appropriately interpret the data. In addition to information on the source of the data, when and why they were collected, who they represent, and limitations of the data, visualizations should include narrative text or other data that put outcomes in context and illuminate the systems that create disparities.

> **Visualizing data in context**
>
> A [2020 ProPublica interactive report](#) titled What Coronavirus Job Losses Reveal about Racism in America allowed readers to explore trends in employment outcomes by race, gender, age, education, and income. As users scroll down the page, they see subgroup comparisons in employment trends. Narrative text in callout boxes provides structural interpretations for the shown disparities. Rather than exclude or combine subgroups with very small sample sizes (for example, Native American men without a high school degree), the ProPublica team displayed a callout box acknowledging the missing data. At the bottom of the page, text cautions readers against comparing subgroups with small differences and discusses other possible explanations for the trends. By providing contextual information and clearly acknowledging the shortcomings of the data, this data visualization tool offered readers key information to make informed inferences.

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Planning	Build a team with diverse lived experiences to decrease the likelihood that implicit bias might appear in data visualizations. Establish common language norms, review processes, and iterative collaboration at the outset to ensure data teams embed inclusiveness in their own processes and therefore, their products.
Analysis	Acknowledge whom the analysis or resulting visualization does *not* represent. Acknowledging which groups are missing, whether due to insufficient data or the focus of the study, leaves space for improvement in future efforts. Consider whom to include in the "other" category and whether such a category is necessary. Identify the contextual information needed to appropriately interpret the data, including any limitations.
Reporting	Ensure visualizations are accessible and are not likely to cause harm, such as by reinforcing stereotypes (consult the Urban Institute's Do No Harm Guide[1528] for specific guidance on colors, labels, ordering, graphics, and icons). Provide opportunity for feedback, allowing community members to validate or reject the narrative portrayed and confirm that the visualization is easy to interpret. Although receiving feedback from community members is not always possible, try to offer them access before publication.

Reflection questions

- Which groups or findings are readers' eyes drawn to in this visualization? Is that the focus of the analysis?

- What does the ordering or spatial organization of the data imply, even if inadvertently?

- Do the colors, pictures, or icons reinforce any stereotypes? Could this visualization cause any potential harm if interpreted incorrectly?

- Which groups are considered in the "other" category? Do they exhibit similar trends, or are you grouping them for convenience? Can you use another term instead?

- Is the visualization's message clear and easy to interpret, without requiring large amounts of text? If not, is a visualization necessary?

Be on the lookout ⚠️

Be careful to not consistently place one race or gender as the default group in visualizations. Across U.S. government surveys and data reports, including the census, "White" is listed first and coded with a "1" in data records. Using "White" as the default or the primary group in data visualizations suggests that the experience of White people represents the benchmark, or standard, to measure desired outcomes against. Altering the order in which data appear depending on the focus of the analysis can not only avoid perpetuating harmful norms, but can also convey findings more clearly and meaningfully.

Additional resources

- [Do No Harm Guide](). This comprehensive guide by the Data Quality Campaign offers principles, norms, and pitfalls to consider when applying equity awareness in data visualization. It includes a racial equity in data visualization checklist to keep on hand when producing data visuals.

- [Reverse Engineering Data Viz for Equity](). This We All Count article details how data users can test their data visualizations against an audience's understanding by using the Reverse Legend test. This technique helps assess how accessible a graphic is or how clear its message comes across to broad audiences if taken out of context.

- [Designing Data Visualization with Empathy](). This article by Bui argues for an empathy-centered approach to data visualization. The author highlights the focus of human-centered and person-first data use, arguing that focusing on the individual behind the data point through graphics, narrative, and context leads to stronger action.

PRINCIPLE 7: Restore communities as data experts using culturally responsive approaches to engagement and co-creation that support equitable data use.

Inequitable power dynamics between data users and communities can perpetuate the disparities that data users aim to address. However, these power dynamics are not inevitable: data users can and should proactively mitigate unintended consequences by involving communities in all phases of the data life cycle, from planning through co-creating solutions. Intentional engagement can promote mutual understanding of assets and challenges within a community, ensuring that data projects are relevant to communities, and that results can be used to drive meaningful change. Restoring communities as data experts involves more than simply offering a seat at the table. It means creating roles for community members to meaningfully impact or lead decision making, valuing their expertise as an integral part of the process, and building relationships rooted in respect to bridge data, policy, and practice.

Data users should seek to understand which communities are affected, both directly and indirectly, by the issue being addressed. In the context of E-W systems, community members might include students, families, educators, and more. Data users should further consider identifying which groups are adversely affected through an intersectional lens, such as Black students with disabilities. Then, data users should identify ways to embed community perspectives *throughout* the project, starting with its conception. Single, point-in-time engagement is typically insufficient—isolated outreach after decisions have been made may be seen as a "box-checking" exercise to nominally gather input. For example, rather than facilitate a single community listening session, data users might recruit community members with relevant lived experience for a recurring advisory council. In its most robust form, this might take the form of CBPR, in which community members actively engage as equal partners in the data project.[1529] However, no engagement model is one-size-fits-all, and community members might play a variety of roles depending on the project's scope, purpose, and timeline.[xxxiii] Building in multiple entry points and avenues for engagement or feedback is essential.

Communities, especially marginalized communities, are often burdened with data initiatives that extract information for personal and institutional gain. To build trusting and productive relationships, data users should define clear roles and expectations for engagement, while collaborating with community partners to determine preferred engagement methods (for example, is it more feasible for community members to participate virtually or in person? During the workday or in the evening? Would they prefer to provide written or verbal feedback?) and opportunities to reduce barriers to participation (for example, by providing child care for in-person activities). Community members should also be equitably compensated to ensure that the partnership is mutually beneficial, and to signal that community members' time and expertise are valued at levels commensurate with that of other experts. Data users should look for opportunities to build capacity within the community as part of the engagement (for example, through collaborative learning processes for data analysis and interpretation) to promote the community's ability to advocate for itself and drive sustained progress beyond the conclusion of the data project. Engaging community members and co-creating opportunities to honor their expert knowledge are foundational activities to successfully implement all data equity principles described in this report.

[xxxiii] See Methods and Emerging Strategies to Engage People with Lived Experience (Skelton-Wilson et al., 2020) for a discussion of various roles for individuals with lived experience, including storyteller, advisor, grantee, partner, or staff member.

> ### Community collaboration in NYC improves student outcomes
>
> In New York City's Community Schools model, the district provides formal support for data sharing and collaboration between school leaders and community partners. Confidential data-sharing agreements enable schools *and* communities to access secure, real-time data on attendance, behavior, and course performance. School leaders and community partners meet regularly to review data, interpret trends, and identify appropriate interventions. The city's Office of Community Schools provides training and support on meeting facilitation, which includes guidance related to inclusive decision making. A study by the RAND Corporation showed that within the first three years (2015–2018), community schools positively affected attendance, on-time grade progression, and high school credit accumulation, while reducing rates of chronic absence. Other state and district education leaders can apply lessons from New York City to promote meaningful community participation in decision making ([Data Quality Campaign, 2018](#)).

Applying this principle throughout the data life cycle

Key phases for this principle	Example applications
Context-setting	Identify what you mean by "priority communities," that is, who is directly and indirectly affected by the focal issue. Be careful not to assume that racial, ethnic, or socioeconomic diversity indicates lived experience relevant to the project. Collaborate with community members to align on what the key issues are and which perspectives to prioritize. Examine potential power dynamics between data users and communities.
Planning	Recruit members of priority communities to participate in initiative teams or advisory councils. Honor the intersectionality of collaborators' identities by recruiting individuals who have had a variety of experiences within the same community and therefore might bring nuanced perspectives on the issue or project. Establish decision-making criteria that systematically incorporate community perspectives. Use facilitation methods that promote equitable participation. For example, if facilitating a meeting involving policymakers and community partners, design activities that capture equally weighted input from all participants, such as anonymous ranked-choice voting.
Analysis	Add dimension to findings through anecdotal and contextual information from lived experiences. Engage community partners when reviewing preliminary findings to validate that data have not been misinterpreted.
Reporting	Visualize and communicate data and findings using plain language so that they are easy to interpret, accessible to communities, and can be used to drive change. Share data in a variety of formats, such as at town halls, at cultural events, and via email or webinar. Build trust with communities by providing timely access to data. For example, if a school administration is evaluating whether to include a program in its budget for the next school year, the administration must receive information before the budget is due to support data-driven decision making.

Reflection questions

- Which groups would this data project affect? Who can help validate our understanding of key groups or illuminate blind spots?

- Who can we recruit from priority communities to participate throughout the project life cycle? How will we reach them? How will we compensate them for their involvement?

- How will we systematically incorporate different groups' perspectives in decision making?

- What has the community engagement process revealed about the experiences, burdens, and benefits for different groups?

Be on the lookout ⚠️

Be careful not to exploit or tokenize lived experience. Feeling pressure to speak on behalf of an entire community can be burdensome for people. Avoid suggesting a monolithic view of "community" by incorporating a variety of perspectives and honoring the diversity of experiences within communities. For example, invite several members from the community with diverse backgrounds to serve on an advisory council, not just a single representative. To avoid exploiting lived experience, data users should also take an inclusive, human-centered, trauma-informed approach to engaging the community to mitigate the risk of retraumatizing individuals when discussing potentially sensitive topics.

Additional resources

- [Why Am I Always Being Researched?](). This Chicago Beyond resource offers practical guidance for community organizations, researchers, and funders looking to address inequities and unintended bias in research projects.

- [Methods and Emerging Strategies to Engage People with Lived Experience](). This brief by Skelton-Wilson et al. discusses strategies and best practices for engaging people with lived experience in federal research initiatives and discusses how they may serve in various roles.

- [Making Racial Equity Real in Research](). This report by Creger, geared toward funders, researchers, and community partners, offers five key steps to establishing effective partnerships using an anti-racist approach.

- [Engaging People with Lived Experience Toolkit](). This step-by-step guide, developed by 100 Million Healthier Lives, includes supporting resources and examples to help data users effectively and equitably engage with community members with lived experience.

- [The Spectrum of Community Engagement to Ownership.]() This toolkit by Facilitating Power helps data users understand and apply a spectrum of community partnership models, ranging from consultation to community ownership.

Data equity principles endnotes

[1510] Hill, H. C. (2016). *50 years ago, one report introduced Americans to the black-white achievement gap. Here's what we've learned since.* Chalkbeat. https://www.chalkbeat.org/2016/7/13/21103280/50-years-ago-one-report-introduced-americans-to-the-black-white-achievement-gap-here-s-what-we-ve-le

[1511] Faria, A. M., Sorenson, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). *Getting students on track to graduation: Impacts of the Early Warning Intervention and Monitoring System after one year.* Regional Educational Laboratory West, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/rel/Project/388

[1512] Evgeniou, T., Hardoon, D. R., & Ovchinnikov, A. (2020). *What happens when AI is used to set grades?* Harvard Business Review. https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades

[1513] Strauss, D. H., White, S. A., & Bierer, B. E. (2021). Justice, diversity, and research ethics review. *Science, 371*(6535), 1209–1211. https://doi.org/10.1126/science.abf2170

[1514] Diakopoulos, N., Fridler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., Jagadish, H. V., Unsworth, K., Sahuguet, A., Venkatasubramanian, S., Wilson, C., Yu, C., & Zevebergen, B. (2017). *Principles for accountable algorithms and a social impact statement for algorithms.* Fairness, Accountability, and Transparency in Machine Learning. https://www.fatml.org/resources/principles-for-accountable-algorithms

[1515] Staats, C. (2014). *Implicit racial bias and school discipline disparities: Exploring the connection.* Kirwan Institute, Ohio State Unviersity. https://kirwaninstitute.osu.edu/research/implicit-bias-school-discipline

[1516] Capatosto, K. (2017). Foretelling the future: A critical perspective on the use of predictive analytics in child welfare. Kirwan Institute, Ohio State University. http://kirwaninstitute.osu.edu/wp-content/uploads/2017/05/ki-predictive-analytics.pdf

[1517] U.S. Department of Education. (n.d.). *What is FERPA?* Protecting Student Privacy. https://studentprivacy.ed.gov/faq/what-ferpa

[1518] GDPR.EU. (2018). *What is GDPR, the EU's new data protection law?* Proton Technologies AG. https://gdpr.eu/what-is-gdpr/

[1519] Clark, J. (2020). *How does the CCPA impact education?* Edlink. https://ed.link/community/what-is-ccpa/

[1520] Mid-Atlantic Equity Consortium (2021). *Equity audit.* https://maec.org/wp-content/uploads/2021/02/Preview-MAEC-Equity-Audit-2021.pdf

[1521] District of Columbia Office of the State Superintendent of Education. (2019). *OSSE section 2 analyzing root causes.* Archived Teacher Equity Guidance and Resources. https://osse.dc.gov/node/1552426

[1522] American Society for Quality. (n.d.). *Fishbone (ishikawa) diagram.* https://asq.org/quality-resources/fishbone

[1523] Lawless, K. (2013) *Root cause analysis tree diagram.* Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Root_Cause_Analysis_Tree_Diagram.jpg

[1524] Andrews, K., Parekh, J., & Peckoo, S. (2019). *How to embed a racial and ethnic equity perspective in research.* Child Trends. https://www.childtrends.org/publications/a-guide-to-incorporating-a-racial-and-ethnic-equity-perspective-throughout-the-research-process

[1525] Moon, T. R. (2011). *Preliminary information.* Project Implicit, Harvard University. https://implicit.harvard.edu/implicit/takeatest.html

[1526] Chicago Beyond. (2018). *Why am I always being researched?* (p. 69). https://chicagobeyond.org/researchequity/

[1527] General Services Administration. (n.d.) *Section508.gov.* https://implicit.harvard.edu/implicit/takeatest.html

[1528] Schwabish, J., & Feng, A. (2021). *Do no harm guide.* Urban Institute. https://www.urban.org/research/publication/do-no-harm-guide-applying-equity-awareness-data-visualization

[1529] See Andrews et al. (2019).