DATA JUSTICE GUIDEBOOK

PRODUCED BY
ADRIEL NEELY, MPH, MBA
REVISIONS BY CHANDLER WYATT

UNITED WAY
Southeastern
Michigan

Version 2

Table of Contents

Purpose and Goal of the Data Justice Guidebook

As an organization we are committed to deconstructing systemic racism and institutional bias while promoting Diversity, Equity and Inclusion (DEI) as a core value in every aspect of our work. United Way for Southeastern Michigan believes data can help tell the story of our region. Our analysts and visualization designers strive to make data accessible, meaningful, and actionable for all, regardless of comfort or experience. Find more information about our data visualization philosophy here.

The goal of the Data Justice Guidebook is to provide guidance and resources for how to approach the data life cycle through a DEI mindset. It offers a set of recommendations and tools to center equity and justice when sharing, using and analyzing data that can help move us toward our Objective.

This guidebook encompasses all aspects of data, ranging from data collection, design of visualization, and analysis. For each guidebook section there is a brief explanation of the resulting topic, examples on how to address DEI concepts within data work and specific tools to be intentional about the work we do here at UWSEM.

United Way for Southeastern Michigan (UWSEM) is an institution that has relied heavily on data to address social issues in the four-county area. Our region is over-surveyed, over-surveilled, and over-studied. UWSEM is not exempt from playing a role in this system. This guidebook offers an opportunity to ensure that we, as an organization, look at how we approach data collectively and apply a DEI mindset to our data work.

What is Data Justice?

Data justice is an approach that remedies ways of collecting and disseminating data that have harmed historically marginalized communities. For centuries, data has been weaponized against Black, Indigenous, and people of color (BIPOC) communities and other marginalized communities; has been used to reinforce oppressive systems that result in divestment and inappropriate, harmful policies.

The collection of and processing of massive amounts of data has become more commonplace. Trust in government, research institutions, social organizations, and academia is at an all-time low. Data sharing and integration across sectors has enabled the transformation of aggregate and individual-level data into information that can be used to address immediate and long-term needs of the community. These data have also been used to improve social services and develop innovative interventions, with the ultimate goal of building stronger communities. However, in the same way data are being used to strengthen communities, these data are also used to reinforce racist polices, exacerbate inequities, and harm communities they were designed to protect. University of Pennsylvania researchers, Amy L. Hawn Nelson and Sharon Zanati, in the International Journal of Population Data Science, argue that those traditionally overrepresented within data systems (mainly people living in poverty and BIPOC individuals), are negatively impacted by laws, policies and business rules that are reinforced by the data collected. Marginalized groups can be denied services, access to opportunities, or wrongfully exploited, impacting the ability to participate fully in society.

With the newfound impetus on utilizing interconnectedness of individual level data, there is a need for a serious critique and level setting with regards to data and data justice. Data justice, as presented in this guidebook, is intended to advance the research of the relationship between data, social justice, and society. Data justice aims to capture lived experiences that are community-driven, centered to counter harm and systemic erasure perpetrated on BIPOC communities vis-à-vis oppressive data practices (both intentional and unintentional).

While research has long tried to determine the essential pillars of data justice, the fundamental premises of data justice are that data should: (1) protect individuals from harm (2) make visible community-driven needs, challenges, and strengths, (3) be representative of community; and (4) be used in ways that promote community self-determination.

Centering Equity Throughout the Data Life Cycle

Cross-sector data sharing and integration has transformed how we understand and address long-term community needs. However, the data collected are used (intentionally or unintentionally) to reinforce racist, ableist, sexist, and harmful practices that produce inequitable access and outcomes.

The first step to centering equity in data is to acknowledge the history and negative implications of data for marginalized groups in our current social and political system. We must be intentional on how we center data justice in the Data Life Cycle; work to use data to reduce harm and improve outcomes for all. We know that centering equity and justice in our approach to data is not binary. It is an iterative process with small steps toward a more equitable approach to data. There are numerous ways to center equity and justice across the stages of the data life cycle; this guidebook can provide a foundation for that growth. Each stage offers an opportunity to incorporate equity and justice into the organizational data framework.

Guidebook Format

Definition: Present and define the stage of the data life cycle.

Guiding Questions: Provide suggestions/recommendations on how to address data justice within that stage of the life cycle.

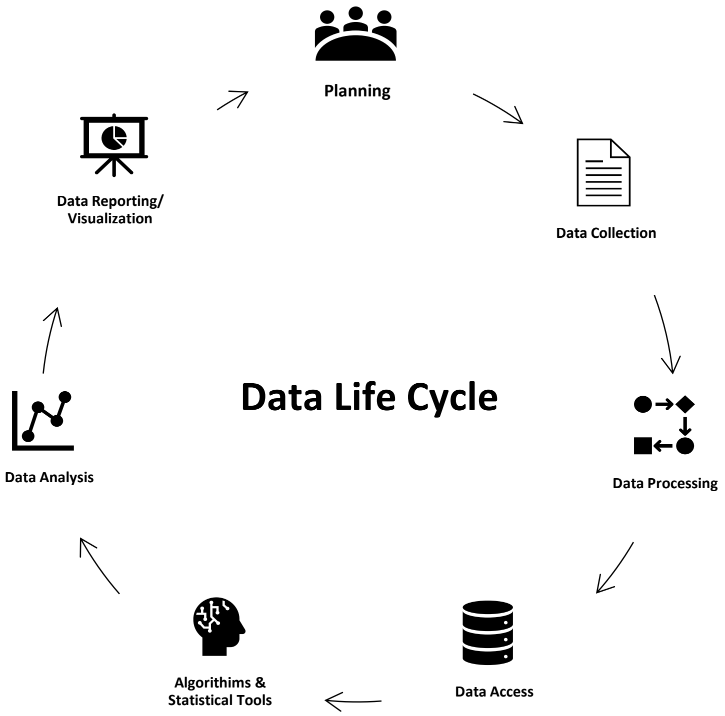
Positive Data Practices: Present positive data practices.

Equity in Action: Provide a real-world example of an organization putting data justice techniques into practice.

What is the Data Life Cycle?

Simply stated, the data life cycle is the entire period of time that any data exists in your “system.” Different data will go through different stages (often more than once) at different times. Traditionally, lessons and insights learned from one step in the data life cycle will inform future data steps and projects, creating a loop.

There is an opportunity to center data justice and equity in each stage of the data life cycle.



Stage 1: Planning

Definition

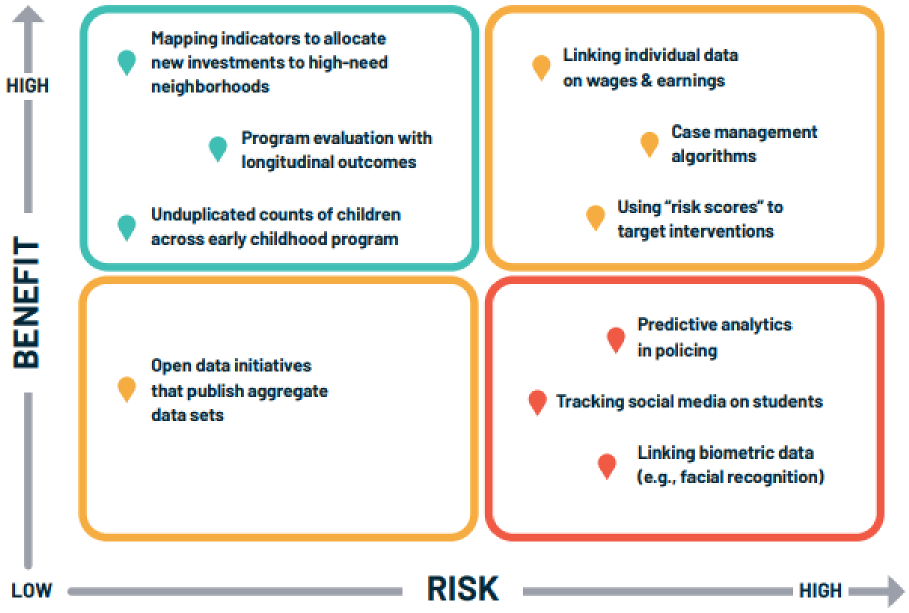
Planning is the first stage of the data life cycle, and it includes all of the work to prepare for future stages of the data life cycle. This includes, but is not limited to, articulating the purpose and mission for the data, understanding the racial, social, and historical context of the data, creating ethical guidelines, developing a project plan, and convening appropriate work groups. Bringing a racial equity mindset to the planning stage will help create a standard for incorporating racial equity in the data life cycle.

We suggest that teams build the following four questions into their planning process (especially when data is being generated, collected, and/or analyzed).

Guiding Questions

- ✓ **Why is this work necessary?**
    - If this initiative does not advance organizational strategy through responsible data practices, including the use of Artificial Intelligence (AI), it may warrant pausing at the planning stage until alignment is clear. Projects should demonstrate measurable impact, transparency, and compliance with emerging AI governance standards.
- ✓ **Who does this work benefit?**
    - . If we cannot identify intended beneficiaries such as communities, stakeholders, and end-users while ensuring equal outcomes, the project should be paused and re-centered.
- ✓ **How does it benefit the community at large?**
    - If this data project offers no real, tangible benefit to the community (or that benefit is unclear) conversations should be had as to the necessity of the product.
- ✓ **Who can the data process harm?**
    - If the full lifecycle of the data reveals harm to marginalized or vulnerable groups, such as privacy breaches or surveillance risks, it is our duty to pause and implement strategies to mitigate this risk. This includes bias audits *(e.g., checking for higher error rates for marginalized groups)*, fairness checks *(e.g., comparing outcomes across demographic groups)*, and privacy-preserving techniques *(e.g., anonymizing or aggregating sensitive data)*.

These steps will be instrumental in creating a project plan, identifying stakeholders, and creating an equity framework throughout the data life cycle. The image below provides a guide on how to look at these four questions.



If you find that a potential project falls into a high-risk/low-benefit box, high-risk/high-benefit, or low-risk/low benefit category, it may warrant a pause of the current work. A revaluation of the data needs and a possible re-organization of the work is needed to ensure that the project moves closer to a high-benefit/medium-low risk category.

Positive Data Practices

Including diverse perspectives (such as community members with lived experience and agency staff who understand the data) on planning committees.

Building capacity for researchers, administrators, and community participants to work together on agenda-setting.

Researching, understanding, and disseminating the history of local policies, systems, and structures involved, including past harms and future opportunities.

Building data literacy among organizations and community members, which could range from light engagement through public activities like data "gallery walks" to more intense involvement, such as community-based participatory action research.

Establishing a shared data language and agreed upon standards for reporting community indicators, aligned with open data, frameworks, sources and methods for reporting on community-based indicators.

Clarifying decision making authority on problem framing and question setting.

Elevating community-driven research priorities to funders, shaping responsible AI and data funding strategies that focus on equity and sustainability.

Equity in Action: Data Planning

Tacoma Equity Index

The City of Tacoma, Washington, created a Strategic Visioning Framework to guide the city's decision making through 2025. Collaborating with community members and groups during this planning process helped center themes of equity, opportunity, partnerships, and accountability in Tacoma's strategic vision. As a result, Tacoma established both the Equity and Empowerment Initiative and the Office of Equity and Human Rights to operationalize the city's strategic vision for service delivery around inclusivity and accessibility. Working with the Kirwan Institute of Race and Social Justice at The Ohio State University, the City of Tacoma created an Equity Index to measure social mobility in the city. Drawing from the concept of opportunity mapping, the Equity Index uses 20 indicators—aligned with the city's strategic goals—to measure opportunity and equity in the city. To date, the tool has been used to inform community conversations on equity and disparities and to guide policy makers' approaches to decision making and resource allocation.[3]

Stage 2: Data Collection

Definition

Data collection, the process of gathering information to inform the study of a policy, program, or problem is a critically important step in the data life cycle. It includes all types of data collected by an organization including administrative data, participant intake data, self-reported online data, service payment records, and participant surveys data used to evaluate programs or services. How we collect data reflects our organizational values. Our data collection methods can perpetuate bias, inequity, and exclusion in the same way they can promote equity and justice within our communities. In the following section, we will focus on best practices for collecting demographic data, outline the benefit of disaggregated data to support data collection, and offer recommendations for designing an inclusive survey.

Guiding Questions

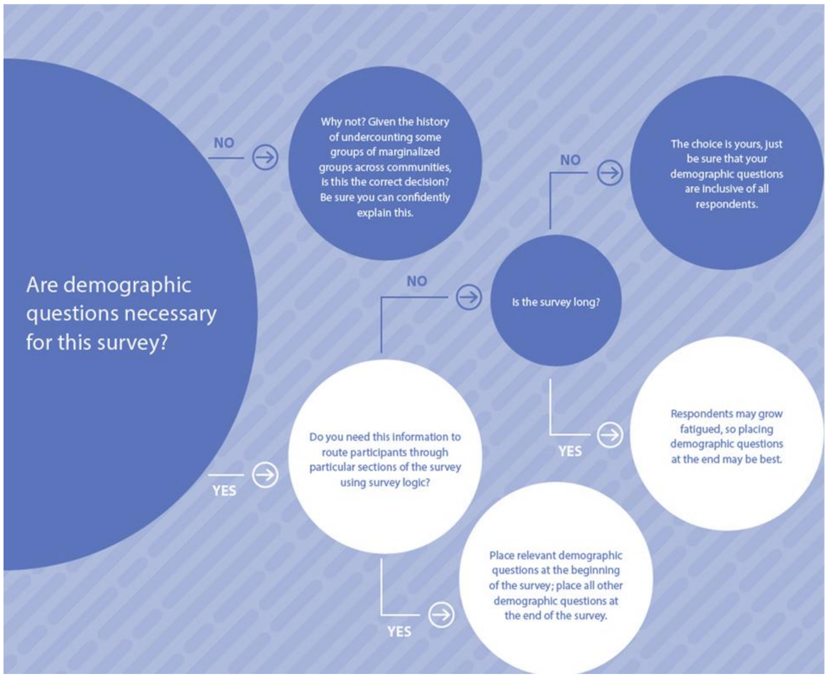
Before determining the content of demographic data collection, we should work to decide what demographic data to collect, how to approach data privacy and where to place demographic questions within the tool. These suggestions are adapted from the Charles and Lynn Schusterman Family Philanthropies "Guide toward Diversity, Equity and Inclusion in Demographic Data Collection" and we encourage teams to go through the following checklist before embarking on demographic data collection of any kind:

- ✓ **What is the purpose of my demographic data collection tool? (e.g., RSVP form, application, program evaluation), and is demographic data collection absolutely necessary?**
- ✓ **Does the program or strategy reach a specific population? What is that population?**
- ✓ **What are the criteria for distinguishing between "nice to know" and "need to know" data? What specific decisions will the data help inform? Will the data be pertinent and actionable?**
- ✓ **Who will review the demographic data? Who will use these data to make decisions? Are we collecting this data ethically and transparently?**

If the answer to one or more of these questions is ambiguous, it would be beneficial to pause and re-evaluate the data collection method and/or seek alternatives.

The following guiding questions offer a guide to demographic survey data collection and survey design.

1. **Where should I place demographic questions?**
   - After going through the aforementioned checklist and deciding if you will move forward with collecting demographic data, it is important to consider the placement of demographic questions in your tool. Placing demographic questions in particular places can affect bias. We recommend that demographic questions be placed at the end of a survey to limit response bias and reduce survey fatigue. Below is a flow chart that can be used to determine the most effective place for demographic questions.[4]



2. **How do I ask for demographic data?**
    - ***Solicit feedback from the communities that will be responding***. Time and resource permitting ask for input from the communities that represent your population of interest; engaging feedback from individuals that you serve is a good practice to integrate into organizational culture.
    - ***Include a statement about consent and confidentiality***. Obtaining informed consent is a basic ethical obligation and a legal requirement for almost all data projects.

Example Statement

Thank you for taking the time to participate in ____ Survey. The purpose of this survey is to…. [Explain]. Please note that participation is voluntary and all answers and comments you provide will remain 100% confidential. No one from our program team will have access to your responses, only [Department/Individual] will have access to the data, and we will safely secure the data. If you have any questions before participating or need special accommodations to participate, please contact us at [Enter Email].

Before you begin the survey, please select an answer to the below questions to provide electronic consent.

Have you read the above information?	Yes/No
Do you voluntarily agree to participate in this survey?	Yes/No

- ***Explain why you are asking for demographic information***. Respondents want to know how UWSEM will use their data, how it benefits them and how we will protect their information. You can apply this principle to any personal information you collect from survey respondents. It is especially important to protect the data of individuals from marginalized communities. Explaining how you will use the data to support, rather than harm underserved communities can reduce understandable fear.

- ***Provide multi-select checkboxes or open-ended questions for surveys when possible***. It is best practice to use either of the two options or a combination of the two.

  Multiselect: Identity is complex, the best approach for inclusive data collection is to avoid giving respondents only one possible answer. Allow respondents to select multiple answers to give them more freedom to express their diversity.

  Open-Ended: In open-ended questions, respondents can clearly express their view of themselves and use the terms they feel are appropriate to describe their identities. Additionally, collecting data on how people self-identify in open-ended questions over a long period can inform the wording of future survey questions.

Example of Multiselect and Open-Ended Question Combination

I identify as: *(Select all that apply)*

☐ Asian (East Asian, South Asian, Southeast Asian)
☐ Black or African Descent
☐ Hispanic, Latino/a/x, or Spanish Origin
☐ Middle Eastern or North African
☐ Multiracial or Multiethnic
☐ Native Hawaiian or Other Pacific Islander
☐ Native or Indigenous People (American Indian, Alaska Native, First Nations, Aboriginal, etc.)
☐ Prefer not to answer
☐ Prefer to Self-Describe here: ________________

- ***Include "Prefer not to answer" and replace "Other" response choices*****.** Consider making questions optional or including a ***"prefer not to answer"*** option. It is also important to include a response choice that has an option for individuals to enter information that does not appear on a list. However, having to enter something in an "Other" category can marginalize and alienate respondents. Instead, use language like ***"Prefer to Describe"*** or ***"Prefer to Self-Describe"***.

- ***Assess the order of response choices. Display options (especially race and ethnicity) alphabetically****.* The order in which response choices appear can reinforce implicit bias. Display options (especially race and ethnicity) alphabetically or manually arrange choices to support an inclusive survey experience.

3. **How should I approach other demographic question options?**
   - *Gender Identity*: As an organization, we must ensure that demographic survey questions about gender identity and sexual orientation represent how individuals perceive themselves, and that all individuals feel seen and heard. Gender identity varies by culture so the guidance included here must be tailored to communities in question. As a general recommendation, we suggest that a user collecting gender identity use an open-ended description with a pronoun multiselect list.

Example Gender Identity Questions

How do you currently describe your gender identity? *(Select all that apply)*

- ☐ Man
- ☐ Woman
- ☐ Cisgender
- ☐ Agender
- ☐ Non-binary
- ☐ Transgender
- ☐ Transfeminine
- ☐ Transmasculine
- ☐ Two-spirit
- ☐ Prefer not to answer
- ☐ Prefer to self-describe: _________________

How do you currently describe your gender identity? *(Open-ended question)*

____
_

Please note, it is best practice to not ask pronouns unless you are going to be addressing someone using those pronouns. Pronouns are not necessarily a proxy for gender identity. For more resources on sexual orientation and gender identity definitions, please visit information presented by the *Human Rights Campaign*: Sexual Orientation and Gender Identity Definitions.

Sexual Orientation:

Example Sexual Orientation Question

How would you describe your sexual identity? *(Select all that apply)*

- ☐ Asexual
- ☐ Bisexual
- ☐ Gay
- ☐ Heterosexual/straight
- ☐ Lesbian
- ☐ Pansexual
- ☐ Queer
- ☐ Prefer to self-describe: _____________
- ☐ Prefer not to answer

- Please note that the choices above are listed in alphabetical order. Over time, analyze responses to see what self-describe options are more frequent and use that information to inform edit the survey question.
- It is recommended to list choices in alphabetical order. For more definitions on sexual orientation, please visit: LGBTQ Terminology.
- ***Race and Ethnicity*****:** For questions about race and ethnicity, it is best practice to:
  - **Allow the respondents to select all the choices that apply, rather than forcing them to choose one.**
  - **Ask race/ethnicity in one question rather than two.**

  Extensive research and usability testing has shown that separate race and ethnicity questions are confusing and undercount Hispanic respondents, while a combined race and ethnicity question with detailed checkboxes is the optimal design for collecting accurate race and ethnicity data. As with other demographic questions, it is best practice to list choices in alphabetical order.

Example Race/Ethnicity Question

Which of the following best describes you? *(Select all that apply)*

☐ Black or African American
☐ American Indian or Alaska Native
☐ Asian or Asian American
☐ Hispanic, Latinx-a/o or Spanish Origin
☐ Middle Eastern or North African
☐ Native Hawaiian
☐ Pacific Islander
☐ White or Caucasian
☐ Not listed here or prefer to self-describe: ________________
☐ Prefer not to answer

- ***Disability***: Asking questions about disability is a delicate topic so we encourage you to visit resources listed below when discussing disability in a respectful and inclusive manner:
  - I am Disabled: On Identity-First Versus People-First Language by Cara Leibowitz
  - Guidelines for Writing about People with Disabilities by the Americans with Disabilities Act National Network
  - Disability Language Style Guide by National Center on Disability and Journalism
  - Accessibility and Etiquette Resources by RespectAbility
  - Jisc: Creating Accessible Surveys This resource provides considerations for survey design and technical delivery.
  - University of California, Office of the President: Guidelines This guideline page provides helpful tips for creating surveys that are accessible and easy for all to navigate.

Positive Data Practices

Adhering to data management best practices to secure data as they are collected—specifically, with carefully considered, role-based access. This is a collaborative effort that involves IT. It is best practice, as an organization, to limit who has access to personal, identifiable information. IT is essential in creating role-based access/security.

Including agency staff and community stakeholders in defining which data should be collected or reused.

Collaborating to develop a shared data collection agenda that is connected to practice, policy, and research. This data collection agenda should be communicated at the organizational level.

Collaborating with agencies and community to generate a data development agenda—a plan for access and use of data that are needed to answer high-interest questions (e.g., expanding gender identity categories on a registration form; building support for digitizing eviction records).

Working with staff to support equity-oriented data collection practices (e.g., programmatic staff to update a registration form, technical staff to update a "forced" field on a data entry platform).

Recommendations for specific demographic categories (i.e. race, gender identity) for surveys that we administer.

Emphasize the importance of qualitative data. Encourage the use of interviews, open-ended responses, focus groups, and community listening sessions to capture context that quantitative data cannot provide around identity, lived experience, and barriers to access.

Working with and developing flexible data systems that adapt to context, environment, or system changes.

Identify and analyze reasons for opting out of demographic questions. Use patterns in opt-outs to refine wording, assess trust issues, reduce sensitive data collection when not needed, and design safer, more transparent practices.

Consider digital access when collecting data and understanding history behind digital divide and data collection.

Equity in Action: Data Collection

Allegheny County, PA, Department of Human Services by Shauna Lucadama & Jessica Ruffin

In 2009, after being approached by community members, Pennsylvania's Allegheny County Department of Human Services (DHS) began a department-wide effort to look at ways to provide culturally responsive, affirming services to the LGBTQ communities involved with DHS. At the time, gender and sexuality were not openly discussed in regard to DHS practices. In 2013, through a partnership with The Center for the Study of Social Policy, DHS began the work of piloting guidelines for collecting data related to sexual orientation, gender identity, and gender expression (SOGIE) for children and youth involved with the child welfare system. The steering committee created to lead this effort ultimately recommended updating the department's case management system to include SOGIE data, implement standards of practice related to SOGIE, and provide training and support to case workers implementing these new practices. Before collecting new data, DHS had to address privacy and data security concerns surrounding youth SOGIE data, the implications of data being shared with external stakeholders, and the complexities and costs of updating an IT system. Additionally, DHS engaged with IT staff to ensure they knew the importance of these inclusive changes in order to mitigate any harm during the design process. A field test of the guidelines showed the broader child welfare community the importance of SOGIE data in improving services to the LGBTQ communities. Now, these guidelines are standard practices in providing services to families.

Stage 3: Data Processing

Definition

Once data has been collected, it must be **processed**. There are many processes involved in cleaning and preparing raw data for later analysis. While the order of operations may vary, data preparation typically includes integrating data from multiple sources, validating data, and applying transformations. Data is often reformatted, summarized, standardized, and enriched as part of the data process workflow. Ethical data processing emphasizes accountability, reproducibility, and equity safeguards; especially where cleaning decisions can influence outcomes of the data.

Guiding Questions

In order to increase transparency and ease of replicability, we recommend development of a data process summary. This is a cover page or an end page, on all products, with a statement of data transformations and/or cleaning methods. This extends to explaining why decisions were made, who contributed, and limitations. By doing so, we can recall the steps that were taken to process data and why we, as a team, came to that decision.

As public institutions increasingly rely on data and AI-supported decision-making, processing workflows should not only prepare data for analysis but also remain transparent, participatory, and accountable to the communities reflected in that data.

With specific reference to data cleaning and transformation, not all data cleaning techniques and practices are created equal. There is a degree of manual intervention that is usually required to provide an initial first check. The following are recommended data cleaning techniques:

- ✓ **Drop Irrelevant Data**
  - Identify and eliminate data that may skew the outcome of your current project. Clarify that the data you want to remove is not essential to the dataset. Erasing meaningful data can skew results, creating bias and inaccuracy in the results of the data.
- ✓ **Get Rid of Duplicate Data**
  - Data cleaning would help to spot any such duplication. You would then merge the records or remove one of the duplicates. Any two records with the same unique ID in a database table or file are duplicates.
- ✓ **Structural Errors/Discrepancies**
  - There are different types of structural errors, from typos to inconsistent capitalization. These can be a problem in categorical data or grouped datasets, so they need cleaning. Consider using bar plots to spot structural or capitalization errors in your datasets. These graphs show data categories along one axis and their corresponding values or characteristics on the other. For example, "asphalt" and "Asphalt" may appear separately on a bar plot despite being the same type of roofing material.[14]
- ✓ **Address Outliers**
  - Any value that substantially differ from the rest of your data may be an outlier. Before removing data, consider community or demographic context. Certain outliers may represent marginalized experiences that should be preserved rather than discarded.
- ✓ **Drop or Flag Missing Data**
  - You may drop observations that have missing values when analyzing statistical data. In this case, dropping is better than imputing values that can impact computation results. You may fill missing numeric data with "0", however, statistical calculations should ignore these zeros.
- ✓ **Standardize the Data**
  - Cleaning your data includes standardizing it to have a uniform format for each value. Aligning text, dates, numeric values and consistent naming formats are ways you can clean and standardize your data by giving it uniform. You may start by having all strings in the same case (upper or lower). Metric conversion may be necessary when standardizing measurements.[14]
- ✓ **Validate the Data**
  - Validation ensures your data is correct and ready for meaningful analysis. So, check all your data for accuracy after changing it. Critical considerations in the final stages of data cleansing include ensuring that: Your data meets pre-established range constraints, there are no missing values for mandatory fields and there are no nonsensical values.[14]

For more techniques and suggestions on data cleaning please visit: The Ultimate Guide to Data Cleaning.

Positive Data Practices

It is important to adjust data cautiously, retaining information that is essential for answering questions or telling meaningful stories, while recognizing that removing data deemed “irrelevant” can unintentionally skew results or weaken the integrity of the analysis.

Building capacity for researchers, administrators, and community participants to work together on agenda-setting.

Collaborating with end-users to address outliers in your dataset. This work should be collaborative and people-centered if possible. Outliers may represent marginalized experiences, not errors every time they occur.
Having due diligence and exploring all the potential differential impacts of data cleaning across demographic groups and bias equity findings.
Working with staff to support equity-oriented data validation practices (e.g., ensuring data meets pre-established constraints, ensuring there are no missing values for mandatory fields, making sure there are no inconsistences in how identity is recorded)
Treat equitable data cleaning as an iterative and time-intensive process; not a quick technical step.
When possible, use scripts and data logs to track changes.
When using AI-based cleaning, balance automation with human oversight to eliminate AI-bias and ensure transparency.

Equity in Action: Data Processing

The MAPS Project

The MAPS Project is a student-centric initiative to chart the rapidly evolving higher education system and works to bring high-quality data and historically marginalized voices to higher ed decision-makers. The MAPS Project aims to Model, Analyze, Prototype, and Share innovative solutions to challenges in higher education. With COVID-19 forcing the closure of campuses early this spring, colleges and universities moved classes online and closed housing facilities, separating students from their campus communities and resources. This disruption accelerated several trends that already existed in higher education, with a disproportionate impact on students from historically marginalized backgrounds. Given the rapid change brought on by COVID-19 and augmented by the United States' accelerated fight for racial justice, it is more important than ever that students' voices and experiences are included in conversations on the future of higher education. To begin this work, the Sorenson Impact Center MAPS project convened students from across the nation to share experiences, perspectives, and needs from their institutions. To support this need, the MAPS Project through inclusive data cleaning processes, created the Financial Health Indicators Index (Financial Health and Equity Dashboard) to leverage six years of data and sector-specific models as an interactive resource to provide decision makers with a data-driven way to understand the financial state of all students, especially the underserved and historically underrepresented/marginalized.

Stage 4: Data Access

Definition

Data access refers to practices and policies on who can securely obtain, view, and use data, when it may be accessed, and for what purpose. Organizations should strive to use clear, multi-tiered approach to data access that begins with a clear delineation of practical and legal data availability.[3] The determination of which data is open, restricted, or unavailable can have significant implications for equity. For example, the intended or unintended consequences of a data release may disproportionately impact some individuals or communities more than others. Alternatively, the impact of not releasing data that is needed to understand or address a community-based problem may be disproportionately felt. Responsible access to data requires a balance of transparency, privacy, security, and equity.

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

Guiding Questions

Categorizing data as open, restricted, or unavailable is a large task but one that is important for centering equity. Both restricting and opening access to data can lead to equity concerns. The risks and benefits should be carefully considered.

When utilizing restricted data in an ethical manner it is important to understand:

- ✓ **Why do you want to access the information?** (e.g., research, evaluation, or another purpose)
- ✓ **What type of information do you want to access?** (e.g., aggregate information or identified records)
- ✓ **Who is the audience?** (e.g., agency analyst, independent evaluator, or concerned citizen)
- ✓ **How will you share the information?** (e.g., anonymized findings via dashboard, aggregate reporting in a static document, publicly released names).[3]
- ✓ **Have the risk and benefits of dataset have been weighed?** (e.g., service improvements, privacy exposure).

Positive Data Practices

Open data that have been identified as valuable through engagement with individuals represented within the data.

Clear data release schedules and information on where to go and how to access data once they are released.

Utilize transparent data request procedures and explain how requests are evaluated.

Adhering to data management best practices for data access, e.g., secure transfer of data (encryption in transit and at rest); clear data destruction parameters, if applicable, following use
Utmost care given to de-identification and anonymization of data prior to release (Data Suppression/De-Identification Guidelines)[32]
Accessible data request process with clear policies and procedures for submitting a request and how requests are evaluated
Creating, using, and sharing high-quality metadata to inform requestors about what data are available.
Free or low cost for data request, with costs clearly communicated (e.g., based on hourly rate or a flat fee)
Supporting directory-level data sharing among agencies for referral coordination and to lessen administrative burden on clients
Clear documentation of why data are unavailable (e.g., specific statute, legislation, data quality explanation, data are not digitized, undue burden in data preparation)

Equity in Action: Access

Birth through Eight Strategy for Tulsa (BEST) /Asemio17 Privacy-preserving Record Linkage Pilot by Jessica England & Dan Sterba

In 2019, eight organizations entered into a formal data collaborative in Tulsa, Oklahoma, aiming to identify and address community-level challenges regarding race, equity, and service overlap. The collaborative, known as the Birth through Eight Strategy for Tulsa (BEST), includes 12 years of data from county government, local government, nonprofit, private sector, and philanthropic organizations that represent 32 programs and 220,000 unique individuals. BEST piloted a new platform developed by Asemio17 utilizing privacy-preserving record linkage that supports data integration while keeping individual and organizational data private and secure. The platform's use of cryptographic and record linkage technology for community analytics allows researchers to integrate data more quickly, at lower cost, while enhancing privacy for individuals and organizations. The process, driven by agency leaders incentivized by access to more robust information and improved service delivery, produced results in less than two months while improving ethical and security protections.

Stage 5: Use of Algorithms & Statistical Tools

Definition

Automated or semi-automated algorithms are, in computer programming terms, a set of well-defined instructions to solve a particular problem. They are designed to take a set of inputs and produce the desired output.[33] Research and practice consistently demonstrate that algorithms are not value-neutral; they reflect historical inequities, institutional priorities, data limitations, and design choices made by people. As a result, algorithmic systems can reproduce or amplify structural bias if not intentionally governed. Responsible use of algorithms requires proactive transparency, continuous bias assessment, documented limitations, and evaluation of potential social impacts across affected populations.

Guiding Questions

Our suggestion for formalizing a commitment to data justice and increase transparency is to draft a social impact statement that describes how the five principles of algorithm/machine learning (responsibility, explainability, accuracy, auditability, and fairness) will be operationalized in practice.[34] This statement will explain the intention to responsibly create the algorithm, explain the algorithm in simple terms, highlight the accuracy of the algorithm, acknowledge that our algorithms should be auditable, and ensure that our algorithms are as fair as possible (not deliberately introducing bias and harmful data practices).

Positive Data Practices

Involving diverse stakeholders, including specific community advisory boards, in early conversations about the algorithm design, procurement, or implementation decisions..

Determining responsibility for oversight of algorithm development and implementation, with clear communication channels for input.

Mandatory impact assessments that involve thoroughly thinking through potential intended and unintended consequences, including equity, privacy, and downstream effects across communities.

Clearly communicating anticipated benefits, limitations, and risks of algorithmic tools to internal and external stakeholders using accessible language.

Human-led algorithm use, where humans retain authority to review, override, pause, or discontinue algorithmic outputs—particularly in high-impact or high-risk contexts.

Transparency regarding what data drive the algorithm and how, e.g., description of design and testing process, list of factors that the tool uses, thresholds used, outcome data used to develop and validate the tool, definitions of what an instrument forecasts and for what time period.

Efforts to improve the quality of data included within the algorithm, including efforts to balance the use of risk and protective factors.

Using “early warning” indicators to provide meaningful services and supports to clients.

Evaluating algorithms using multiple performance and fairness metrics, including disaggregated error rates (e.g., false positives/negatives) across race, gender, disability status, and other relevant characteristics.

Equity in Action: Use of Statistical Algorithms and Tools

Automating NYC by AISP with contributions from Deepra Yusuf, Elyse Voegeli, Akina Younge, & Jon Truong

New York City convened the Automated Decision Systems Task Force to assess the use and proliferation of automated decision systems (ADS) across social and city services in May 2018. The task force, which was mandated under Local Law 49, required participants develop a series of recommendations on the use of ADS and would include community outreach and input as part of this process. In January 2019, four graduate students—Akina Younge, Deepra Yusuf, Elyse Voegeli, and Jon Truong—constructed the website Automating NYC to help make conversations about ADS more accessible to community members before the first public engagement forum hosted by the Task Force in late April. It aimed to support community members in asking informed questions about how automated decisions can and do contribute to unjust systems, and with the hope that future systems are built to benefit them. Members of the group were all previously involved in efforts to promote algorithmic justice and had been accumulating knowledge on NYC-specific ADSs since fall of 2018. To build the website, they used information from investigative reporting, research, and open data, and worked with New York agencies to develop case studies. Accessible language and formatting, as well as a series of examples across different social services systems, allows for a broad audience with varying experience or awareness of ADSs. Nontechnical topics are used to describe technical terms and concepts, such as buying avocados to explain a decision tree algorithmic (adapted from Cathy O’Neil’s book Weapons of Math Destruction), with engaging and interactive activities. Individual stories always accompany technical explanations, thereby merging theoretical explanation of automated decision making with its application and the felt impact on the local population.

Stage 6: Data Analysis

Definition

Data analysis is the stage of the data life cycle at which data are examined, interpreted, and synthesized to generate findings, insights, and conclusions that inform decision-making. Data analysis may range from descriptive summaries, which use simple statistics and charts to describe patterns and trends in the data, to exploratory analysis and more advanced approaches such as statistical modeling, causal inference, and predictive or prescriptive analytics.

In practice, data analysis often occurs at a distance from the communities, environments, and systems represented in the data. When analysis is conducted without contextual knowledge or stakeholder input, it can produce misleading findings and reinforce structural inequities. Incorporating participatory interpretation, contextual expertise, and reflexive analytic practices reduces the risk of harm and improves validity.

Data disaggregation is a critical equity tool, but it presents tradeoffs. Without disaggregation, inequities may be masked and experiences rendered invisible. With disaggregation, analysts risk over-surveillance, stigmatization, or misuse of subgroup data. Navigating these tensions requires intentional, iterative, and collaborative analysis grounded in data justice principles.

Guiding Questions

Analysts must confront our own assumptions and implicit biases that influence our analysis and subsequent visualization. We must:

- ✓ **Be aware of cultural differences among the priority population.**
- ✓ **Ensure that there is diversity among evaluators/analysts.**
- ✓ **Highlight shared background/life experiences with the priority population.**
- ✓ **Engage in cultural humility and bias-awareness training for evaluators.**
- ✓ **Have personal awareness of cultural frameworks, assumptions, and biases.**
- ✓ **Identify conditions or pressures that may trigger analytic bias**

How does data justice factor into the two main types of data analysis?
Data justice informs both quantitative and qualitative analysis by shaping how data are framed, interpreted, contextualized, and acted upon. Equity-centered analysis prioritizes understanding root causes, structural drivers, and lived experiences—rather than reinforcing deficit-based narratives or surface-level comparisons

Quantitative Data Analysis
Equity-centered quantitative analysis requires more than disaggregation by demographic categories. Analysts should examine how systems, policies, and environments shape observed outcomes and avoid treating identity categories as causal explanations. **Data disaggregation** allows staff to examine important variables by different subgroups and to carefully examine the distribution of important variables for the population (racial/ethnic identification, gender identity, sexual orientation, first-generation status, veteran status, diagnosed disabilities or impairments, etc.). Whenever possible, researchers should disaggregate by subgroups to uncover the heterogeneity of experiences both between and within racial and ethnic groups. ***It is extremely important to note that data disaggregation should aim to go beyond racial and/or ethnic group classification to look at structural and social determinants that might explain observed findings.***

We recommend organizations:

- ✓ **Explore the intersectionality of given identity with other dimensions of identity. Ask why the trends revealed in this intersectional analysis may occur.**
  - – For example, when showing graduation rates by school and by racial groups, it may be equally important to show the financial resources provided to each school, the local history of school segregation, or the financial hardship faced by students' families.
- ✓ **When using secondary data sources, check the metadata (when possible) to see who is being left out in analysis**.
  - – For example, are we missing incarcerated populations, populations without internet access and homeless populations?
- ✓ **Use Equity Gap Scores (when possible).**
  - – For example, in Census Report from the City of Detroit it is apparent that the equity gap is increasing.
- ✓ **Pay close attention to comparison groups.**
  - – For example, be careful of the historical context when comparing groups by identity such as comparing white and Black home ownership rates.
- ✓ **Understand that bias is inherent.**
  - – But checking for biases in your hypothesis or null hypothesis can help examine what you consider the norm to be.
- ✓ **Double check analysis projects/models for biases and assumptions.**
- ✓ **Ensure that data analysis is a continuous feedback loop.**
  - – For example, engage in a post-analytic deliverable conversation with the end user and staff to learn from each other to make significant contributions to building a high quality, equitable data culture.

Qualitative Data Analysis

Qualitative analysis centers everyday human experiences and understandings of the world. It shines in its ability to capture the richness and depth of data and is essential for understanding how policies, programs, and systems are experienced by people in real-world settings—especially where quantitative data alone is insufficient or misleading. Qualitative analysis is valuable for situating and interpreting findings in a specific context or capturing how a given issue may be understood from different positions or angles.

Through the collection and analysis of narrative and storytelling, qualitative research offers important perspectives and information not captured by quantitative research methods. We recommend the following steps, when applicable, as we approach applying an equity mindset to qualitative analysis.

- ✓ **Use an Equity Glossary or a Key Equity Terms & Concepts Glossary to develop a codebook**.
  - – These codes, or markers of specific pieces of data, will help frame your overall themes and interpretation. This codebook will be published; it will be the first step in standardizing qualitative analysis (especially when examining the same groups of individuals across data projects).
- ✓ **Honor participant language.**
  - – Communities express constructs and concepts in different ways. Use the community's own language when describing themes and ultimately when reporting. Be aware of your own process of selecting supporting quotes.
- ✓ **Avoid lifting out a stereotypical narrative of deficiency.**

  - – Be sure to lift out strengths and assets of stakeholder groups.
- ✓ **Use narrative to describe themes, allowing intersectionality to play out.**
- ✓ **Collaborate to make meaning of data.**
  - – Whenever possible, budget enough time into the project to allow for a collaborative interpretation of the findings to co-create meaning of them with diverse stakeholders.
- ✓ **Present equity-related limitations to the group.**
- ✓ **Responses should be filtered by themes to notice differences by race, gender, power level and other characteristics, if the interview sample is large enough.**[37]

When conducting any type of data analysis (quantitative or qualitative) staff must remember and keep in mind that:

- ✓ **Communities and populations are not only research subjects.**
  - – We must be respectful of, transparent to, and reciprocal with all communities and populations.
- ✓ **We are not omniscient.**
  - – A community's perspective can complement and supplement the researchers' knowledge with contextual factors.
- ✓ **Community interpretation can surface blind spots, biases, and unintended harms.**
- ✓ **Allocate time and resources for collaboration with diverse stakeholders.**

Positive Data Practices

Using participatory research to bring multiple perspectives to the interpretation of the data (This includes but is not limited to community members and staff input). UWSEM's focus is more internal, by centering staff input into data analytics plans (for example). This could include project specific engagement with different impact teams depending on the questions asked.

Engaging domain experts (e.g., agency staff, caseworkers) and methods experts (e.g., data scientists, statisticians) to ensure that the data model used is appropriate to examine the research questions in local context.

Correlating place to outcomes (e.g., overlaying redlining data to outcomes), contextualizing data and visualizations.

Using appropriate comparison groups to contextualize relationships. Random sampling (targeted) so we are getting voices from minority groups but still get their perspectives (to stay statistically sound, but also inclusive in our sample).

Employing mixed methods approaches when developing the analytic plan, including purposefully seeking out qualitative data (interviews, focus groups, narrative, long-form surveys) in conjunction with quantitative administrative data to better understand the lived experience of clients.

Disaggregating data and analyzing intersectional experiences while applying safeguards against misuse. (i.e., looking at race by gender).

Empowering professionals and community members to use data to improve their work and their communities. Short term: show intent to drive action that benefits those being served. Long-Term: Workshops to involve community voice in all levels of analysis.

Facilitating ongoing dialogue within organizations to clarify what “community voice” means in practice and how it informs action.

Equity in Action: Data Analysis

#ChangeFocusNYC by Sarah Zeller-Berkman

In 2016, the Administration of Children’s Services (ACS) in New York City partnered with the Department of Education to examine a series of indicators common to both departments’ datasets. The project brought together agency leads, community activists, and academics to form a community engagement committee tasked with ensuring that nonagency voices helped shape the analysis. The committee held meetings with individuals working in schools, community-based organizations, and health clinics to identify indicators related to education and child abuse, prevention, and detention that would be analyzed by researchers. Fifteen youth were chosen to partner with academics at the Youth Studies Program at the CUNY School of Professional Studies to design and implement a participatory action research project formally titled #ChangeFocusNYC. #ChangeFocusNYC set out to answer two main questions: What are the experiences of NYC youth ages 14–21 who have dealt with multiple city agencies in their lives? What are policy and/or programmatic recommendations that could benefit youth ages 14–21 who are dealing with multiple agencies in their lives? Youth investigators were involved in all phases of the research project and were essential contributors during development of the analytic plan. Collaboratively generated answers to the research questions will help ACS work toward creating a system in which young people are continuously engaged in shaping the institutions that impact their lives.[3]

Stage 7: Reporting, Visualization and Dissemination

Definition

Reporting, visualization, and dissemination comprise the final stage in the data life cycle, where findings from the data are communicated internally and/or externally. This stage determines how data is interpreted, acted upon, and remembered. This involves not only the audience but the method of dissemination methods. Dissemination methods include briefs, interactive documents, websites, dashboards, social media content, data walks, posters, infographics, and spreadsheets. Across all these mediums, centering equity means paying attention to:

- ✓ **Which data are highlighted?**

- ✓ **How are the findings framed and contextualized?**
- ✓ **Readability and interpretability**
- ✓ **Accessibility across ability, language, and technology.**

Content intended for public audiences should avoid unnecessary technical jargon, use person-centered language, and be translated into languages relevant to the community context. Translation should be viewed as a design requirement—not an afterthought. King County in Washington state offers a to assist users in creating written products in other languages.[23]

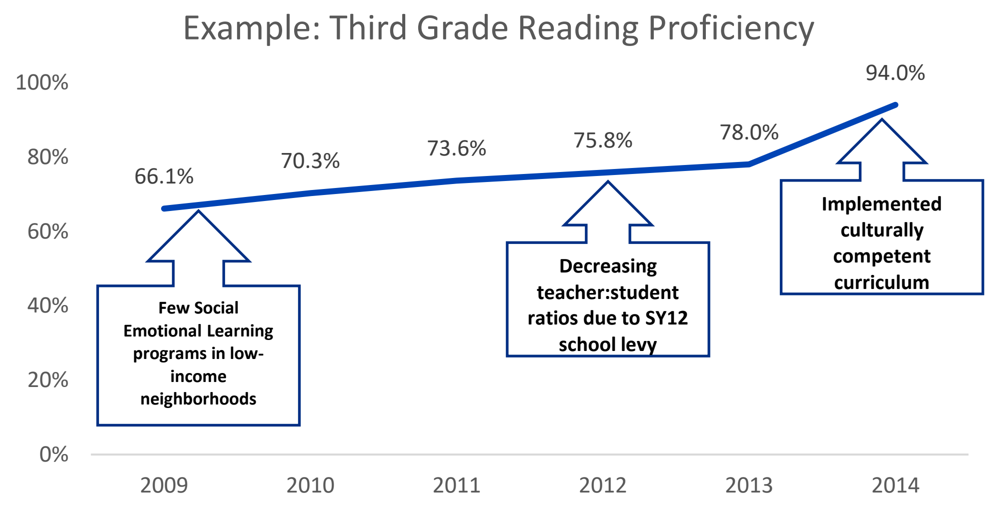
Guiding Questions

- ✓ **Critically examine the data.**
  - Staff should understand where the data come from, who is included/excluded from the data, how these data were collected, and who it benefits and/or harms.
- ✓ **Use language with equity awareness.**
  - Texts, titles and labels are the first things readers observe when encountering a visualization. They should use people-first language.[24] Start with the person, not the characteristic when describing groups of people with labeling. For example, “an individual with epilepsy” is a person-focused phrase, while “an epileptic person” is disability focused. It is important to note that language continues to evolve, so please keep in mind that labels may change with time. To that end staff should consider including a footnote or endnote explaining why a particular term was used.
- ✓ **Order labels purposefully.**
  - Oftentimes our graphs, charts and dashboards include demographic data that is ordered in the way they appear in the raw data. We should take an active role in how to present data from different groups. Which group we choose to highlight first can reinforce a harmful relationship between groups in question. There is no set of rules when ordering data, but staff should consider:
    - **Does your study focus on a particular community? If it does, that group should be presented first.**
    - **Is there a particular argument or story you are trying to tell? If so, the order or presentation of results should reflect that argument.**
    - **Is there a quantitative relationship that can guide how the groups are ordered? Can they be sorted alphabetically or by population size, sample size (weighted or unweighted), or magnitude or effect of the results?**
- ✓ **Acknowledge missing or suppressed groups.**
  - It is extremely important to acknowledge who is and is not included in final visualizations. Many charts on race and ethnicity, for example, only show Black people, Latinx people or white people. Smaller racial or ethnic groups are lost. Oftentimes, this is a result of data limitations and sample size issues. Similarly, gender breakdowns often neglect nonbinary and transgender people. While it is important to avoid including a catch all “Other” category when presenting visualizations, situations do arise that it is okay to use an “Other” category to protect anonymity of groups with small number of respondents. In cases where data were collected about a specific group, but that group was not presented in the visualization, it may be worth listing these decisions in the chart’s notes. This can both inform readers that data do exist for these smaller

groups (even though they may be limited) and offer transparency about the chart maker's decision about groups they included in their visualization.

- ✓ **Consider colors in visualizations.**
  - – Ensure that the color palette is consistent for people with color vision deficiencies, and the contrast between those colors and white/black text meet basic accessibility guidelines. Use a Contrast Checker.[25] Do not use color palettes that reinforce gender or racial stereotypes, such as baby pink and baby blue to represent women and men. Staff should avoid using colors associated with skin tones and light to dark color palettes to represent demographic groups.
- ✓ **Consider icons and shapes.**
  - – We have to be careful and thoughtful when using icons in any data visualization. When showing groups of people, we should consider a mix of genders, races, and ethnicities. When using icons, avoid icons that make inappropriate depictions of people or communities or reinforce stereotypes. Recognize the cultural significance of symbols; avoid discriminatory and racist imagery. Rather use images that show people as empowered.
- ✓ **Reflect lived experiences/Demonstrate Empathy.**
  - – A challenge in visualizing data is the ability to help readers connect with the context of the data. Standard graphs like bar charts, line charts, and pie charts, are informative, but can abstract from the context and people being represented. Taking an empathetic view of the reader's needs as they read or perceive information is one step to better data communication. Viewing empathy through a DEI mindset would mean considering how the specific lived experiences and perspectives of our readers (not to mention the actual study populations) will perceive the information. Connecting directly with people and communities and trying to better understand their lived experience can help content producers create visualizations and tell stories that better reflect the true experiences of different people.
  - – A good example of incorporating lived experiences can be found in the visualization below. The graph examining third grade literacy rates over time has been supplemented with context that may be lost in the original data.

Example of Incorporating Lived Experiences in Visualizations



- ✓ **Consider the needs of your audience.**
  - Make sure the data are presented in a format that is useful to the audience, that the language is easy to understand, and consider translating products into languages used by audiences. Please find resources about accessibility below.
    - **Plain Language Check**: Checklist for Plain Language[26]
    - **Plain Language Thesaurus**: Plain Language Thesaurus for Health Communications[27]
    - **Readability of text**: Automatic Readability Checker[28]

Power BI has accessibility features built into the product, which don't require any configuration from the report author. Use alternative text (alt text) descriptions to describe the appearance and function of visuals to convey meaningful information on a report. Providing alt text ensures that consumers of your report understand what you are trying to communicate with a visual, even if they cannot see the visual, image, shape, or textbox. Find more PowerBI accessibility resources here.[29] Similarly, Alt Text can be added to Word, PPT, Excel, Sway, and webpages.

The medium used to disseminate findings should match the needs of the community and audiences for whom the product is intended. These mediums could take the form of in-person presentations at community events, data walkthroughs or discussions, infographics for display in community buildings or centers, forums, and interviews with local media.

Finally, dissemination must not end with the report of findings to key stakeholders. These findings should be accompanied by recommendations or actionable items for community members that can be used for sustainability planning or to find solutions to the issues that have been identified. This final step should be considered ongoing.

Special Considerations: Land Acknowledgement

A **Land Acknowledgement** is a formal statement that recognizes and respects Native peoples as traditional stewards of lands. The statement highlights the enduring relationship between Native/Indigenous peoples and their traditional territories, while also acknowledging the historical and present-day impacts of colonization, displacement, and systemic inequities. Land acknowledgements also raise awareness about the Indigenous histories, perspectives and experiences that are often suppressed or forgotten. Land acknowledgments do not exist in the past tense or historical context: colonialism is a current ongoing process, and we need to build our mindfulness of our present participation.[35]

Land Acknowledgement Statement

Our aim is to uphold and advocate for and support the advancement of the sovereignty of Michigan's twelve federally-recognized Indian nations and Indigenous communities and individuals who have lived here and live here now.

We operate on occupied territories called Waawiiyaataanong, named by the Anishinaabeg and includes the Three Fires Confederacy of Ojibwe (Chippewa), Odawa (Ottawa), and Bodewatomi (Potawatomi) peoples. We hold to commit to Indigenous communities in Waawiiyaataanong, their elders, both past and present, and future generations.

We recognize that all Indigenous peoples—here and beyond, whose Land has been occupied—have and continue to be forcibly removed from their homelands. This is in solidarity with Black and Indigenous communities who continue to be systemically excluded and oppressed through the ongoing damaging effects of settler colonialism, genocide, and myths of white supremacy.

To find what land you are on, visit Native Land Digital.[36] To find out more about land acknowledgements, visit Native Land Digital: Territory Acknowledgement[37] or The Native Governance Center.[38]

Positive Data Practices

Creating a range of products to communicate findings across a wide variety of audiences via both online and offline methods of dissemination. Dashboards are not the end all be all for data visualization.

Developing differentiated messaging for different audiences that considers the appropriate level of detail and technical jargon, language, length, format, etc.

Reporting data in an actionable form that supports decision-making and resource allocation to improve the lives of those represented in the data (e.g., analyzing food purchase data to identify food deserts and guide development of grocery stores).

Providing public access to aggregate data (e.g., dashboards, routine reports).

Explicitly acknowledge structural inequities embedded in data.

Including stories as a complement to quantitative findings to better contextualize the lived experience represented by the data.

Conducting impact analyses multiple times during the project (e.g., at the beginning, middle, and end). Asking the core question: does this work mitigate, worsen, or ignore existing disparities?

How is information sent out? How is it reaching communities and other stakeholders?

Providing clear documentation of the data analysis process along with analytic files, so that others can reproduce the results.

Transparency of limitations of the data (throughout the lifecycle)

Equity in Action: Data Reporting, Visualization and Dissemination

Vulnerability, Inequity, and COVID-19: A Portrait of the Pandemic in Allegheny County[31]

The health inequities exposed by COVID-19 underscored the importance of collecting race-stratified data to inform local policymakers. For the public health researchers trying to provide that, the pandemic also revealed some major pitfalls, especially about relying on open-source data. Information is almost never neutral: What gets collected, how it is analyzed, reported, contextualized, and used—all that reflects preexisting assumptions and biases. All of these became factors when RAND, the Black Equity Coalition (BEC), and Surgo Ventures collaborated on a tool to report on COVID-19 vulnerability and disparities using publicly available data in Allegheny County, Pennsylvania. The goal was to help decisionmakers identify geographic areas and racial/ethnic populations most at risk of infection and complications from the novel coronavirus.

To investigate these disparities, researchers visualized existing data on COVID-19 cases and testing rates and compared them with defined levels of vulnerability to COVID-19 within the different municipalities and neighborhoods in Allegheny County. They also explored testing access, testing rates, cases, and such health outcomes as hospitalizations and deaths by race to explore the existence of disparities. Finally, they examined physical distancing over time by vulnerability and demographic and socioeconomic factors. Researchers also examined the association of physical distancing with COVID-19 testing and deaths. The results of these analyses create a portrait of communities that are most vulnerable to COVID-19 and that experience racial inequities. The hope is that these analyses inform individual and policymaker efforts to mitigate the spread of COVID-19 as the pandemic continues and provide a model for future tracking of inequities in future pandemic response.[31]

Conclusion

This guidebook is intended to provide a reference of how to approach day-to-day data work using a DEI mindset. Overall, the results of this process underscore the notion that centering data justice & equity throughout the data life cycle is not a single, discrete step, but rather an ongoing process. Each stage presents new opportunities to apply an equity framework; presents new challenges and considerations. Strategic/organizational response, evaluating existing practices and policy, employee training, and building continual data/information sharing across the organization is key to ensuring that data justice is at the forefront of your work.

Appendix A: Data Justice/Equity Product Checklist

Planning

☐ Have we considered the necessity of this work, who it benefits, how it benefits the community at large, and whether it could produce harm?

☐ Do we have a clear plan for embedding data justice and equity across every stage of the data life cycle, including risk mitigation and governance considerations?

☐ Have we sought community input and researched the historical and structural context affecting this community, program, or issue?

☐ Have we clearly defined goals, decision-making authority, and how success will be measured?

Data Collection

☐ What are the criteria for distinguishing between "nice to know" and "need to know" data?

☐ Is the data we are collecting necessary, ethical, and directly tied to actionable decisions?☐ Have we clearly communicated purpose, consent, confidentiality, and how the data will be used?

☐ Are demographic questions inclusive, optional when appropriate, alphabetically ordered, and designed to reduce bias (e.g., multi-select, open-ended, "prefer not to answer")?

☐ Have we considered accessibility, language needs, digital equity, and community trust when designing our collection tools?

Data Processing

☐ Have we documented our data cleaning, transformation, and validation decisions transparently?

☐ Have we evaluated whether cleaning decisions (e.g., removing outliers, handling missing data) could disproportionately impact marginalized groups?

☐ Are our processes reproducible, accountable, and supported by human oversight when automation or AI tools are used?

Data Access

☐ Have we clearly categorized data as open, restricted, or unavailable and weighed risks and benefits of access?

☐ Are privacy, de-identification, and secure handling protocols in place and documented?

☐ Are data access procedures transparent, equitable, and clearly communicated?

Algorithms & Statistical Tools

☐ Have we documented how responsibility, explainability, fairness, and accountability are embedded in any algorithmic or statistical tools used?

☐ Have we evaluated performance and potential bias across demographic groups, including disaggregated error rates where applicable?

☐ Is there human oversight and a clear process for review, override, or reassessment of automated outputs?

Data Analysis

☐ Have we examined our assumptions, cultural context, and potential biases in interpretation?

☐ Have we appropriately disaggregated data and considered intersectional and structural factors influencing outcomes?

☐ Have we engaged relevant stakeholders or community voices in interpreting findings when possible?

☐ Have we clearly documented limitations and potential equity implications of the analysis?

Reporting

☐ Is our language equity-aware, person-centered, and free from deficit framing?

☐ Are visualizations accessible (color contrast, alt text, ordering of groups) and respectful in imagery and presentation?

☐ Have we acknowledged missing, suppressed, or excluded groups transparently?

☐ Are findings contextualized, actionable, and disseminated in formats appropriate for the intended audience?

References

1. WHY RESEARCH & DATA JUSTICE. Coalition of Communities of Color. (2022). Retrieved 3 July 2022, from https://www.coalitioncommunitiescolor.org/-why-research-data-justice

2. Hawn Nelson, A., & Zanati, S. (2020). Framework for Centering Racial Equity Throughout the Administrative Data Life Cycle. International Journal of Population Data Science, 5(3). https://doi.org/10.23889/ijpds.v5i3.1367

3. Hawn Nelson, A., Jenkins, D., Zanati, S., Katz, M., Berkowitz, E., et al. (2020). A Toolkit for Centering Racial Equity Throughout Data Integration. Actionable Intelligence for Social Policy, University of Pennsylvania.

4. Kaplowitz, R., & Laroche, J. (2020). MORE THAN NUMBERS: A Guide Toward Diversity, Equity, and Inclusion (DEI) in Data Collection [Ebook]. Charles and Lynn Schusterman Family Philanthropies. Retrieved 3 July 2022, from deidatacollectionguide (widen.net)

5. Sexual Orientation and Gender Identity Definitions. Human Rights Campaign. (2022). Retrieved 4 July 2022, from https://www.hrc.org/resources/sexual-orientation-and-gender-identity-terminology-and-definitions

6. Young, K. (2020). LGBTQ Terminology. Keshet. Retrieved 4 July 2022, from https://www.keshetonline.org/resources/lgbtq-terminology/

7. Burkhardt, T., Huang, L. A., Kakuyama-Villaber, R., & Pacheco-Applegate, A. (2021). Racial Bias in Data Assessment Tool. Chicago, IL: Chapin Hall at the University of Chicago.

8. Leibowitz, C. (2015, March 12). I am disabled: On identity-first versus people-first language. The Body Is Not an Apology. Retrieved July 3, 2022, from https://dfwhcfoundation.org/wp-content/uploads/2018/10/I-am-Disabled.-On-Identity-First-Versus-People-First-Language-TBISNAA.pdf

9. Guidelines for Writing About People With Disabilities. ADA National Network. (2021, July 7). Retrieved July 3, 2022, from https://adata.org/factsheet/ADANN-writing

10. National Center on Disability and Journalism. NCDJ. (2021). Retrieved July 3, 2022, from https://ncdj.org/style-guide/
11. WebSymphonies. (2022, June 14). Resources: Respectability. Respect Ability. Retrieved July 3, 2022, from https://www.respectability.org/resources/

12. Creating accessible surveys. Online surveys (formerly BOS) Powerful, flexible online surveys. (2022). Retrieved July 3, 2022, from https://www.onlinesurveys.ac.uk/help-support/creating-accessible-surveys/

13. Regents of the University of California. (n.d.). Guidelines for accessible surveys and forms. Retrieved February 3, 2022, from https://www.ucop.edu/electronic-accessibility/web-developers/tools-and-testing/guidelines-accessible-surveys-forms.html

14. Tableau. (2024). Data cleaning: the Benefits and Steps to Creating and Using Clean Data. Tableau Software. https://www.tableau.com/learn/articles/what-is-data-cleaning

15. Elgabry, O. (2019, March 2). The Ultimate Guide to Data Cleaning. Towards Data Science. Retrieved July 3, 2022, from https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4

16. Sorenson Impact Center: David Eccles School of Business. (n.d.). Financial Health and Equity Dashboard. MAPS Project. Retrieved March 12, 2022, from https://www.mapsproject.org/

17. Krause, H. (2019, October 22). Supercharge your Averages with an Equity Gap Score. We All Count. Retrieved July 4, 2022, from https://weallcount.com/2019/10/22/supercharge-your-averages-with-an-equity-gap-score/

18. Racial Equity Tools. (2020). Glossary. Retrieved August 4, 2021, from https://www.racialequitytools.org/glossary

19. CSSP. (2019). Key Equity Terms and Concepts: A Glossary for Shared Understanding. Washington, DC: Center for the Study of Social Policy. Retrieved from https://cssp.org/resource/key-equity-terms-concepts/

20. Fedorowicz, M., Arena, O., & Burrowes, K. (2020). Community Engagement during the COVID-19 Pandemic and Beyond: A Guide for Community-Based Organizations. Urban Institute. Retried March 15, 2022. Available at: UWSEM Data & Dashboard Hub

21. Washington State Department of Health. (n.d.). Community Engagement Guide. Retrieved March 4, 2022, from https://doh.wa.gov

22. Conservation Law Foundation. (2020, July 21). A Participatory Action Research Field Guide from the Healthy Neighborhoods Study. Retrieved May 16, 2022, from https://www.clf.org/wp-content/uploads/2021/01/PAR-Field-Guide.pdf

23. King County. (n.d.). King County Written Language Translation Process Manual. Retrieved July 4, 2022, from https://kingcounty.gov

24. Examples of Person-First Language. (n.d.). https://www.nationaldisabilityinstitute.org/wp-content/uploads/2023/08/people-first-language.pdf

25. WebAIM. (n.d.). Contrast checker. Retrieved June 4, 2022, from https://webaim.org/resources/contrastchecker/

26. plainlanguage.gov. (n.d.). Checklist for Plain Language. Retrieved 2021, from https://www.plainlanguage.gov/resources/checklists/checklist/

27. Centers for Disease Control and Prevention. (2007). Plain Language Thesaurus for Health Communications. Retrieved 2022, from https://stacks.cdc.gov/view/cdc/11500

28. StyleWriter Software. (n.d.). Automatic Readability Checker (Our Free Text Readability Consensus Calculator). Retrieved January 4, 2022, from https://www.readabilityformulas.com/free-readability-formula-tests.php

29. Davidiseminger. (n.d.). Creating Accessible Reports in Power BI - Power BI. Retrieved July 4, 2022, from https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-accessibility-creating-reports#alt-text

30. Urban Institute. (n.d.). Do No Harm Guide, Applying Equity Awareness in Data Visualization: The Racial Equity in Data Visualization Checklist. Retrieved from https://www.urban.org/sites/default/files/2021/06/08/do-no-harm-guide-checklist.pdf

31. Peet, E., & May, L. (2021). Vulnerability, Inequity, and COVID-19: A Portrait of the Pandemic in Allegheny County Tracking Evolving Disparities in Testing, Cases, Hospitalizations, Deaths, and the Ability to Practice Physical Distancing. https://doi.org/10.7249/tla1080-1-v2

32. University of South Carolina. (2018). Guidelines and Methods for De-Identifying Protected Health Information. Retrieved July 22, 2022, from https://www.schealthviz.sc.edu/Data/Sites/1/media/images/USC_IFS_PHIDataDeIdentification_18.pdf

33. What is an algorithm? Programiz. (2022). Retrieved July 22, 2022, from https://www.programiz.com/dsa/algorithm

34. Diakopoulos, N. D. (2018). Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. FAT/ML. Retrieved July 22, 2022, from https://www.fatml.org/resources/principles-for-accountable-algorithms

35. Newman, J. (2019, April 9). Land Acknowledgement. Native American Cultural Programs. Retrieved July 22, 2022, from https://nacp.uconn.edu/land-acknowledgement/

36. and, O. (2026). Guide to Land Acknowledgements. Guide to Land Acknowledgements. https://nai.msu.edu/resources-container/guide-to-land-acknowledgements

37. Andrews, K., Parekh, J., & Peckoo, S. (2019). How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process. Annie E. Casey Foundation. Retrieved August

1, 2022, from https://www.childtrends.org/wp-content/uploads/2019/09/RacialEthnicEquityPerspective_ChildTrends_October2019.pdf

38. Jones, A. (2021, October 7). Territory acknowledgement. Native Land Digital. https://native-land.ca/resources/territory-acknowledgement/

39. Porter, J. M., Bradford, Z., & Le, L. (2025). The urgent need for data equity, justice, and sovereignty. Stanford Social Innovation Review. https://doi.org/10.48558/50TR-F569

40. University of Washington, School of Aquatic and Fishery Sciences. (2022). Indigenous Peoples terminology guide (Revised April 29, 2022). https://fish.uw.edu/wp-content/uploads/sites/29/2023/08/Indigenous-Peoples-Terminology-Guide-2022-FINAL.pdf

41. Lamsam, T. T. (2025, May 21). American Indian vs Native American: Which term is right and when? Buffalo’s Fire. https://www.buffalosfire.com/american-indian-vs-native-american