# Equity Data Handbook

CA Water Board's Office of Information Management and Analysis (OIMA)

2024-07-01

# Table of contents

# Welcome!

This Equity Data Handbook is an online resource written by the The State Water Resources Control Board (State Water Board) and the nine Regional Water Quality Control Boards (Regional Water Boards), collectively known as the California Water Boards (Water Boards).

Content in this Handbook includes best practices and guidance for Water Boards staff on incorporating racial equity concepts into their data-related work. Specifically, this Handbook will provide guidance and resources to help Water Boards staff conduct each phase of the data life cycle through a racial equity lens - from the planning and design of a project to data collection methods, visualization development (e.g., maps, fact sheets, etc.) and more!

A depiction of the Data Life Cycle is provided below for context. We must have meaningful engagement and partnership with our diverse communities during each phase of the Data Life Cycle, especially those that have been historically underserved namely Black, Indigenous, and other People of Color (BIPOC). If we focus on uplifting those most highly impacted we will inevitably improve the the data products and services we develop and, most importantly, the experiences and outcomes for all communities.

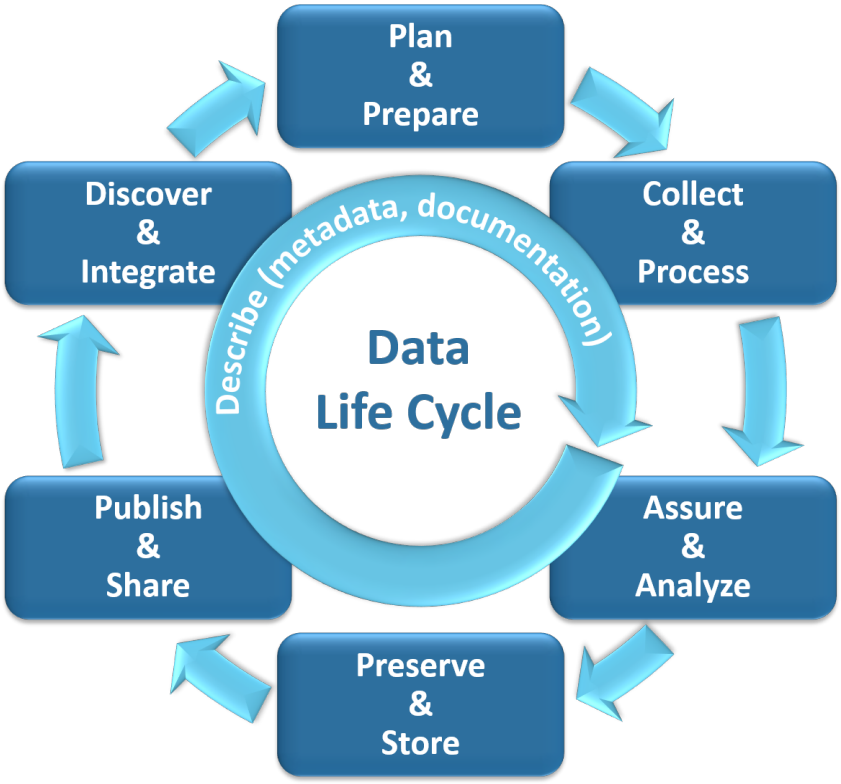


Figure 1: Graphic of the data life cycle.

This Quarto book is an open, living, and continuously iterating resource. If you have suggestions for additions or revisions you think should be incorporated into this book, please email equitydatahelp@waterboards.ca.gov.

# 1 Background

During its August 18, 2020 meeting, the State Water Resources Control Board (State Water Board) publicly acknowledged that the historical effects of institutional racism must be confronted throughout government, and it directed staff to develop a priority plan of action (Aug 18, 2020 Meeting Agenda, Recording). The State Water Board's Racial Equity Team held public and employee listening sessions to help develop a draft resolution. After a public comment period on the draft resolution in spring 2021, the Racial Equity Team made significant updates to the resolution. On November 16, 2021, the State Water Board adopted Resolution No. 2021-0050, "Condemning Racism, Xenophobia, Bigotry, and Racial Injustice and Strengthening Commitment to Racial Equity, Diversity, Inclusion, Access, and Anti-Racism" (Racial Equity Resolution) which affirms the State Water Board's commitment to racial equity and directs staff to undertake a variety of actions to achieve racial equity throughout all State Water Board programs and activities (Nov 16, 2021 Meeting Agenda, Recording).

The Racial Equity Resolution is one milestone on our ongoing journey to operationalize equity throughout our organization. The next step is to implement the Racial Equity Action Plan, which includes specific actions the State Water Board will take to address racial inequities, as well as metrics to measure our progress. With this Action Plan, we envision a sustainable California where race no longer predicts where clean water is available or who has access to it.

> **Note**
>
> It's important to note that racial equity and equity in general is an outcome and there is no such thing as a "racial equity data-set" or "racial equity data".
> Instead we should think of data as a tool to help us achieve the overall outcome of equity - and how we view, use, and act on data and related tools is what will determine whether we operationalize equity or perpetuate injustice.

Development of the Racial Equity Action Plan began in Spring 2022 and involved public and employee engagement and tribal consultations. The Water Boards' Racial Equity Team presented the Racial Equity Action Plan (2023-2025) to the State Water Board as an informational item at the Board Meeting on January 18, 2023 (Agenda, Slides, Recording).

Goal 1a of the Racial Equity Action Plan is to ensure Water Boards data are accessible, equitable, and culturally relevant. One action captured under that goal is the development of a Racial Equity Data Action Plan which must:

1. Develop training and best practices guidance for Water Boards staff on incorporating racial equity concepts into the planning and design of data collection methods and visualizations (e.g., maps, factsheets, etc.) projects.
2. Identify and expand existing opportunities for public participation in science and community data gathering programs to develop new data collection methods, support existing programs, and incorporate community datasets into the database.
3. Create a publicly accessible data catalog tool / interface that includes existing demographic data, Water Boards program data, and other available data (such as heat maps or flood hazard maps) to inform the implementation of the Racial Equity Action Plan.

The Racial Equity Data Action Plan is being developed by the Water Board's Racial Equity Data Subcommittee of the Environmental Justice Roundtable which is a group of volunteer staff from across the Water Board led by the Office of Information Management and Analysis and the Office of Public Participation. This Handbook is intended to address Item 1 above by providing staff a best practices guide for incorporating racial equity concepts into the Water Board programs using data and information. Item 2 above will be iterative and grow as staff and programs begin to utilize the guidance and tools found within this handbook. To fulfill Item 3 above staff have created a California Water Boards Racial Equity Data Resource Hub which will grow as more programs create and publish racial equity based tools and visualizations.

# Part I

# Getting Started

# 2 Getting Started

Tip

**Before embarking on your data project using a racial equity lens you should:**

1. Consult the Racial Equity Data Subcommittee by emailing equitydatahelp@waterboards.ca.gov to help identify audience and potential questions or information your teams' analysis may answer and provide guidance.
2. Review the guidance and best practices described in this Equity Data Handbook and complete the Racial Equity Data Project Form
3. Review the GARE Framework: Normalize, Organize, and Operationalize
4. If possible, take the Advancing Racial Equity training series offered by the Water Boards Training Academy

## 2.1 Consult the Racial Equity Data Subcommittee

Program staff that are beginning data project using a racial equity lens should meet with the Racial Equity Data Team to discuss key aspects of the project including who the audience is and what questions or information you are trying to convey using the available racial equity data. *This engagement should occur prior to beginning any data collection or analysis.*

To initiate a consultation with the Racial Equity Data Team, please send an email to equitydatahelp@waterboards.ca.gov.

## 2.2 Review the Equity Data Handbook and complete the Racial Equity Data Project Form

This Equity Data Handbook is a curated compilation of emerging and comprehensive (but not exhaustive) guidance on:

- How to break down the management questions where racial equity information is being posed against administrative data; and
- How to apply equity best practices during each phase of the data life cycle to begin an iterative process into advancing racial equity.

This Handbook is specifically structured to support Water Boards staff on incorporating racial equity concepts into all phases of their data projects. The guidance and best practices provided serves as a strategic guide emphasizing the importance of collection, analysis and utilization of racial equity data.

These intended users of this Handbook include, but are not limited to:

- Program Staff
- Program managers
- Executives
- Agency Partners
- Tribal Governments
- The Public

After reviewing this Handbook, please complete the Racial Equity Data Project Form so that the Racial Equity Data Subcommittee is better equipped for your project consultation.

## 2.3 Review the GARE Framework

It's strongly suggested staff review the framework outlined by the Government Alliance for Racial Equity (GARE) to normalize, organize, and operationalize racial equity throughout data integration (see image below).

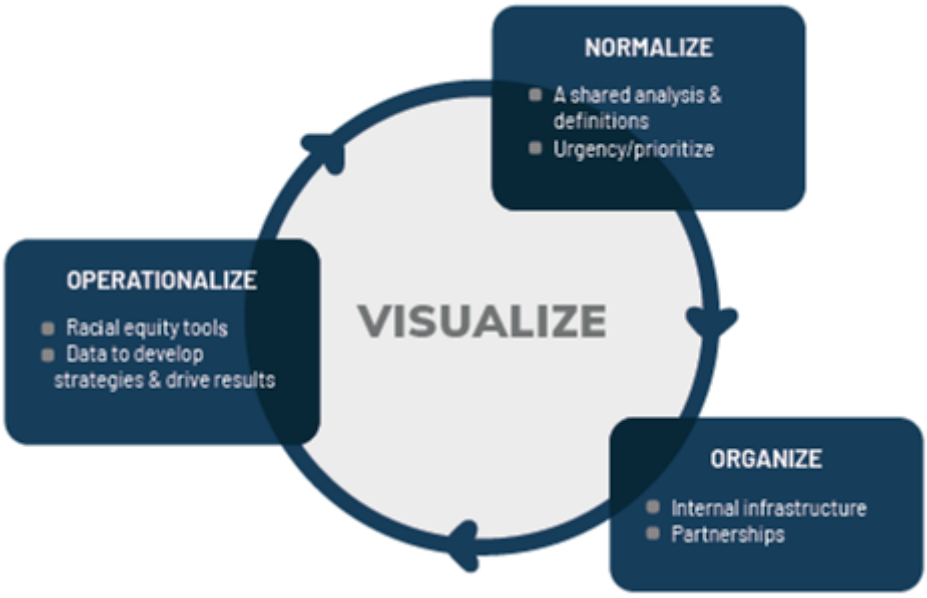


Figure 2.1: GARE model of change. Source: GARE Communications Guide, May 2018

In addition staff should also review why it is important to lead with race:

> "with the recognition that the creation and perpetuation of racial inequities has been baked into government, and that racial inequities across all indicators for success are deep and pervasive. We also know that other groups of people are still marginalized, including based on gender, sexual orientation, ability and age, to name but a few. Focusing on racial equity provides the opportunity to introduce a framework, tools and resources that can also be applied to other areas of marginalization.It is critical to address all areas of marginalization, and an institutional approach is necessary across the board. As the local and regional government deepens its ability to eliminate racial inequity, it will be better equipped to transform systems and institutions impacting other marginalized groups."

## 2.4 Take the Advancing Racial Equity Training

When possible, staff should take the Advancing Racial Equity training series offered by the Water Boards Training Academy to foster a consistent baseline knowledge of racial equity work and the importance of applying a racial equity lens to our work.

# 3 Best Practices

We strongly encourage staff to read this Handbook in it's entirety. However, for those who want to jump to best practices, or who just need a refresher, we have provided them below, with links to pages that have more details, guidance, and resources.

## 3.1 Plan & Prepare

This phase involves conducting an equity assessment (Planning) and developing your data management plan using an equity lens (Data Preparation). Best practices for this phase include:

- **Not rushing through this phase!** We often want to dive straight into "doing something" and taking the time to plan can feel like a waste of time. If done quickly and without using an equity lens, it will be time wasted. Instead, invest the time needed to do this phase well, rather than fast.
- Using this phase to begin the process of trust and relationship building with Tribes, communities, and other expert partners interested in your project. Maybe it involves some early outreach, putting together an email list, establishing a technical advisory committee, or otherwise *co-creating* your planning documents with the communities that will be most impacted by the project's implementation.
- Make the products developed during this phase as open, transparent, and accessible as possible and appropriate. This doesn't necessarily mean that every single thing should be made public - but it does mean seriously considering what can be made open, to whom, and when.
- Use the best available science while relying on current, generally accepted Agency procedures for conducting risk assessments, economic or other technical analysis. **But also understand that just because something has "always been done this way" does not automatically make it the best science, method, or process available.** Be open to integrating other ways of knowing that might be different from "western science" or our business as usual methods or processes, such as the traditional knowledge landscape, and non-tribal expertise that stems from lived experiences.

## 3.2 Collect & Process

This phase involves collecting the data and information you need for your project (Data Collection) and making it tidy so that is ready to be used in your analyses or product development steps (Data Processing). Best practices for this phase include:

- Use already existing and available frameworks and data, supplementing as appropriate.
- Carefully select and justify the choice of the geographic unit you will use in your analysis and discuss any particular challenges or potential aggregation issues related to the choice of spatial scale.
- Keep data disaggregated so you are able to reveal important spatial differences during your analysis phase (e.g., demographic information for each facility/place) when feasible and appropriate.
- Prepare to invest the time required to tidy the data needed for your project. It's commonly understood in the data science field that 80% of data analysis is spent on the process of cleaning and preparing the data (Dasu and Johnson 2003) - expect the same will be true for your project and prepare to invest time and resources accordingly.

## 3.3 Assure & Analyze

Remember - this phase is all about…. Best practices for this phase include:

- Use the highest quality and most recent data available.
- Carefully select and justify the choice of a comparison population group.
- Analyze and compare effects in baseline and across policy scenarios to show differences in effects.
- When data allow, characterize the distribution of risks, exposures, or outcomes within each population group, instead of presenting only average effects.
- Present summary metrics for relevant population groups of concern as well as the comparison population group.
- Be consistent with the basic assumptions underlying other parts of the analysis, such as using the same baseline and option scenarios.

## 3.4 Preserve & Store

Remember - this phase is all about…. Best practices for this phase include:

## 3.5 Publish & Share

Remember - this phase is all about…. Best practices for this phase include:

## 3.6 Discover & Integrate

Remember - this phase is all about…. Best practices for this phase include:

## 3.7 Describe

Remember - this phase is all about…. Best practices for this phase include:

- Discuss the overall quality and main limitations of the data (e.g., completeness, accuracy, validation).
- Discuss available evidence of factors that may make population groups of concern more vulnerable to adverse effects (e.g., unique pathways; cumulative exposure from multiple stressors; and behavioral, biological, or environmental factors that increase susceptibility).
- Identify unique considerations for subsistence populations when relevant.
- Discuss the severity and nature of the health consequences for which differences between population groups have been analyzed.
- Clearly describe data sources, assumptions, analytic techniques, and results.
- Discuss key sources of uncertainty or potential biases in the data (e.g., sample size, using proximity as a surrogate for exposure) and how they may influence results.
- When possible, conduct sensitivity analysis for key assumptions or parameters that may affect findings.
- Make elements of Environmental Justice (EJ) assessments as straightforward and easy for the public to understand as possible.

# 4 Establishing Common Language

When working with a racial equity lens we suggest establishing a common language and definitions to cultivate a collective understanding of underlying concepts and historical context. Creating and agreeing upon a common language can help foster transparency, challenge assumptions, and center the voices of marginalized communities; yet the efficiency of these efforts hinges on a shared language that facilitates understanding and collaboration. By grounding discussions in a common language, we can build trust and empower our team and community.

Establishing a common language and definitions are critical to creating a shared understanding, however we acknowledge that language can be used deliberately to engage and support community anti-racism coalitions and initiatives, or to inflame and divide them. It is important to note that although the language in this Handbook may be commonly used, the list of terms herein is not exhaustive, and may not be the sole definition of a term, and some may disagree with the definitions and their use. More specifically, in this resource we intentionally use the acronym BIPoC (Black, Indigenous, People of Color) as a term that seeks to recognize the unique experience of Black and Indigenous People within the United States. We recognize that naming is power, and we remain committed to using language that supports pro-Blackness and Native visibility, while dismantling white supremacy.

## 4.1 Glossary

Below are a set of key terms and definitions provided by the Water Board Racial Equity Team in the development of the Racial Equity Resolution and Racial Equity Action Plan and are those that we adhere to in this document. (citations can be found here). For a more comprehensive list of equity-related terms, see the Racial Equity Tools Glossary

- **Equality** describes circumstances in which each individual or group is given the same or equal treatment, including the same resources, opportunities, and support. However, because different individuals or groups have different histories, needs, and circumstances, they do not have equal positions in society or starting points. Providing the same resources, support, or treatment does not guarantee that everyone will have fair or equal outcomes.

- **Ethnicity** is a term used to describe subgroups of a population that share characteristics such as language, values, behavioral patterns, history, and ancestral geographical base. Social scientists often use the terms ethnicity and ethnic group to avoid the perception of biological significance associated with race.
- **Institutional racism** describes the ways in which policies and practices perpetuated by institutions, including governments and private groups, produce different outcomes for different racial groups in a manner that benefits the dominant group. In the United States, institutional racism includes policies that may not mention race but still result in benefiting white people over people of color.
- **Race** is a social construct used to categorize humans into groups based on combinations of shared physical traits such as skin color, hair texture, nose shape, eye shape, or head shape. Although most scientists agree that such groupings lack biological meaning, racial groups continue to have a strong influence over contemporary social relations. Historically in the United States, race has frequently been used to concentrate power with white people and legitimize dominance over non-white people.
- **Racial equity** means Race can no longer be used to predict life outcomes and outcomes for all groups are improved. For example, when we hold income constant, there are still large inequities based on race across multiple indicators for success, including the environment, education, jobs, incarceration, health and housing.
- **Racism** is any prejudice against someone because of their race when systems of power reinforce those views.
- **Structural racism** is the normalization and legitimization of an array of historical, cultural, institutional, and interpersonal dynamics that routinely advantage whites while producing cumulative and chronic adverse outcomes for people of color. Structural racism encompasses the entire system of white domination, diffused, and infused in all aspects of society, including its history, culture, politics, economics, and whole social fabric. Structural racism is more difficult to locate in a particular institution because it involves the reinforcing effects of multiple institutions and cultural norms, past and present, continually reproducing old and producing new forms of racism. Structural racism is the most profound and pervasive form of racism; all other forms of racism emerge from structural racism.
- **Systemic racism** can be said to encompass both institutional and structural racism. Glenn Harris, president of Race Forward, defines systemic racism as "the complex interaction of culture, policy and institutions that holds in place the outcomes we see in our lives." The legacy of systemic racism can be seen in a variety of outcomes affecting people of color, such as housing insecurity, a ten-fold wealth gap between white and Black or Latinx households, a dramatic over-representation of people of color in prison, and disparities in education, health, and exposure to environmental pollution.

## 4.2 Non-Inclusive Terms to Avoid

- **"Brown Bag"** - The term "brown bag" has a historical connotation with creating an exclusive gathering that required attendees to have a lighter skin tone than a brown paper bag to participate and gain access.
  - Alternatives: lunch in, lunch and learn, presentation, seminar
- **"Chief"** - this term is used throughout the Water Boards to indicate positions and job titles. This term is appropriated from the Indigenous Peoples of North America and should be avoided wherever possible.
  - Alternatives: manager, lead, head
- **"Grandfathered in"** - The American South created absurd voting requirements that targeted Black people and made it almost impossible to vote. The name for these requirements is the "Grandfather Clause." They wrote the Amendment in a way to imply the practice was not discriminatory. They created stringent new voter requirements such as literacy tests. These requirements did not apply to people who had voted before 1867. Slaves did not know they were free until June 19, 1865. However, slavery was abolished on January 1, 1863, making it nearly impossible for a person formally kept in captivity to be legally allowed to vote.
  - Alternatives: legacied, exempted, preapproved
- **"Master _____"** - using the term "master" to describe something that is the main or centralized source of information is inappropriate due to the connotations associated with slavery.
  - Alternatives: primary, main
- **"White Paper"** - while this term is widely used to describe an authoritative document, the term has historical implications that evoke negative associations especially with Tribes.
  - Alternatives: Issue paper, briefing document, prospectus

# Part II

# Plan & Prepare

# 5 Planning

When planning for your data project with an equity lens, it's best to conduct a thorough equity assessment, or scoping process. Equity assessments[1] are systematic examinations of available data and expert input on how various groups - especially those facing inequity or disparities - are or likely will be affected by a policy, program, or process. They aim to minimize unintended adverse outcomes and maximize opportunities and positive outcomes.

The steps to conducting an equity assessment are described below.

- Step 0: Project Scoping
- Step 1: Describe the selected program, policy, or process, and populations affected by it
- Step 2. Consider historical, societal, and policy context and drivers of disparities
- Step 3. Collect expert input, including from affected community members
- Step 4. Identify information sources and gaps
- Step 5. Analyze program/policy effects - potential or current - on people and communities
- Step 6. Plan for action and accountability

Although these steps are numbered for clarity, teams should synthesize information from all steps rather than completing them in isolation or one at a time.

## 5.1 Step 0: Project Scoping

**Time frame and level of effort:**

- Plan a detailed schedule that accounts for staff availability, budget, technical assistance needs, data availability, and the need to make decisions in a particular time frame.
- Identify and document risks to the timeline, such as staff availability, or threats to the comprehensiveness of the assessment, such as lack of access to experts. Consider potential ways to mitigate risks.

**Project team:**

[1]The bulk of content on this page has been informed by: Conducting Intensive Equity Assessments of Existing Programs, Policies, and Processes (hhs.gov)

- Define roles for team members and assign responsibilities. Plan to share information at key milestones to synthesize information from different assessment steps.
- Consider how to meaningfully involve experts, including people with lived experience with relevant programs and topics; people in communities affected by the program, policy, or process; staff who work with program participants/beneficiaries; or representatives of other offices. Experts can contribute to the assessment process in several ways, such as suggesting data sources, providing multiple perspectives to inform and enhance the analysis, and developing recommendations for action.

## 5.2 Step 1: Describe the selected program, policy, or process, and populations affected by it

Describe the focus of the assessment to provide a foundation for all members of the assessment team and external partners.

- What is the purpose of the selected program, policy, or process, and what are its goals?
- What are the known successes or challenges in meeting those goals?
- What types of actions or policy levers are involved in the selected program, policy, or process (such as grants, contracts, waivers, guidance to partners, technical assistance, or other actions)?
- Which of these actions will be included in the assessment?
- What general descriptive or performance data can the organization use to describe the program, policy, or process (such as number served, total funds distributed, uptake estimates, or other key outcomes)?
- Are there existing quantifiable performance targets relevant to the focus of the assessment? Provide a brief summary.

Identify and describe populations of interest

- What populations are participating in the program, policy, or process, including program participants/beneficiaries? Consider which characteristics are relevant and of interest, such as race, ethnicity, gender identity, sexual orientation, disability status, income, religion, and rural geography.
- What populations are currently left out, or not participating or benefiting at desired rates or at the same rates as others? What are other disparities related to the selected program, policy, or process that are known at the outset of the assessment?

- What are the information sources for those inequities or disparities? What is the comparison population or reference point for observed disparities? Reference point options include the total population in an area, the national population, the largest group, or a benchmark chosen through a planning process. Whenever possible, try to think critically about this population rather than simply defaulting to comparison populations used in the past.
- How might population groups' identifying characteristics overlap in ways that expose them to relatively greater inequities (known as intersectionality)? What implications does this overlap have for the impacts of the program, policy or process? For example, immigrants who are also LGBTQIA+ might face multiple barriers in accessing a particular program.

## 5.3 Step 2. Consider historical, societal, and policy context and drivers of disparities

Describe the context for observed disparities and the program or policy itself.

- What is the social and cultural history of the populations listed in Step 1 and how does this history shape their current conditions? How does this context play a role in how these populations might perceive, access, or otherwise interact with the program or policy?
- What structural or social drivers of disparities might explain observed disparities? Structural drivers of disparities are governing processes and economic and social policies that distribute power and resources in unfair ways, such as an inequitable distribution of emergency funds to certain communities. Social drivers of disparities are differences in the conditions in which people are born, grow, live, work, and age, such as poverty, employment, housing, environment quality, transportation, food security, and community safety. Differences in these social conditions drive disparities. Although these conditions are also known as social determinants of health, this tool uses a broader term to encompass multiple outcomes, including both health outcomes and other outcomes (e.g., economic outcomes). Thinking through these drivers of disparities is important for placing focus on systems and institutions that need to be changed, and it helps to avoid blaming groups of people for poor outcomes.
- What is known about whether structural, systemic, or institutional racism or structural barriers affect the implementation and outcomes of previous programs or policies? Systemic or institutional racism refers to policies and practices that create or sustain disparate outcomes for persons of different races. An example is redlining, where financial services and other housing-related opportunities were restricted for individuals largely based on their race/ethnicity and originating neighborhoods (see this 2021 Memorandum

for the Secretary of Housing and Urban Development regarding Redressing Our Nation's and the Federal Government's History of Discriminatory Housing Practices and Policies)

## 5.4 Step 3. Collect expert input, including from affected community members

Experts can include former or current program participants/beneficiaries, members of communities affected by the program, policy, or process, staff who work with program participants/beneficiaries or affected communities, subject matter experts such as researchers, or staff in other organizations, among others. Sources of expert input on programs, policies, and processes include listening sessions, surveys, interviews, focus groups, and position papers by experts in the field or advocacy groups.

- How will the assessment team engage experts with lived experience with relevant programs, policies, processes, and/or issues? To what extent can these experts be part of the assessment team?
- How will the assessment team engage other experts in the equity assessment (in addition to potentially involving them in the assessment team)? Which experts will be engaged?
- What individuals or communities have historically been excluded or disempowered in decision making? How can they be included and meaningfully engaged?
- How can the assessment team ensure inclusivity when engaging experts, such as translation services or accommodations for people with disabilities? Will there be different options for sharing input for people with different communication preferences or time or transportation constraints?
- How will the assessment team work to decrease power dynamics and ensure that experts are comfortable providing candid input? How can the team be transparent about how input will be shared and used?
- What methods can the assessment team use to collect input, such as focus groups on participants/beneficiaries' experiences with programs? What are experts' experiences with current programs and policies, and what are their views on the benefits and burdens involved in participating?
- What are experts' perceptions about barriers to participation? Can experts help the assessment team understand whether there are current or potential burdens or barriers that are more severe for certain population groups?

> **Tip**
>
> For more detailed guidance and resources regarding outreach and engagement, see the Practical Guidance Document developed by the Office of Public Participation.

## 5.5 Step 4. Identify information sources and gaps

Consider a variety of qualitative and quantitative information sources to support the assessment, including gray and peer-reviewed literature, organization documents and administrative records, surveys, customer inquiry or complaint information, administrative data, program performance data, key informant interviews, and listening sessions or focus groups. Ideally, equity assessments often include both qualitative and quantitative data. Data sources can include, but should not be limited to, expert views.

- What are the quantitative data sources for the assessment process? Quantitative data such as program, administrative, or survey data shed light on the magnitude and prevalence of an inequity or an opportunity for improvement.
- Are available quantitative data disaggregated by relevant variables, such as race, ethnicity, income, and relevant geographic areas? If not, how can the assessment incorporate data that can help organizations understand or estimate the equity impacts of the program, policy, or process?
- What are the qualitative data sources for the assessment process? Qualitative data such as interview or focus group data increase understanding of context, as well as helping to interpret and understand quantitative data.
- Are there gaps or limitations in the information needed for the assessment? If either qualitative or quantitative data are not available, explain why. If there are gaps, how might the assessment team obtain new or better information, or highlight the need for investments in better data? It is important to describe gaps that might reflect historically overlooked inequities or point to the need for information sources that could be developed in future years.

## 5.6 Step 5. Analyze program/policy effects - potential or current - on people and communities

Drawing on all previous steps in the assessment process, analyze the available data and describe equity-related outcomes of the program, policy, or process. Describe findings with as much specificity as possible.

- What quantitative and qualitative analysis methods did the team use to analyze the available data? Did the team synthesize quantitative and qualitative data to develop a complete picture of current inequities or disparities related to the program, policy, or process?
- What are the assessment team's findings on positive and negative equity-related outcomes of the program, policy, or process? What quantitative and qualitative evidence of inequities exists?
- What evidence is there of inequities in areas such as awareness of programs and benefits, processes and rules, administrative burden, access to services, participation, outcomes, quality, and engagement?
- How do findings change the team's understanding of disparities related to the selected program, policy, or process known at the outset of the assessment?
- What factors might be driving observed inequities or disparities? Are any of those factors potentially caused by the program or policy that is the focus of the assessment?
- Have experts helped the assessment team interpret the available data or validate or refine the initial findings?
- In what ways might the findings be limited due to data gaps or analysis constraints? What findings point to the need for further research?

## 5.7 Step 6. Plan for action and accountability

Develop a detailed plan to address inequities identified in Step 5 within the scope of your program.

- What solutions are needed to resolve observed inequities or disparities, or to address identified drivers of those inequities or disparities? Which solutions are in the program's sphere of authority?
- What are the program's short-term and long-term goals for improvement? Quantify those goals if possible.
- What steps will the program take to accomplish each goal? What coordination, training, information systems changes, business process changes, or other implementation actions are needed?
- Have subject matter experts—including those with lived experience—weighed in on needed solutions, proposed goals, or planned action steps? Are all components of the improvement plan responsive to the needs and cultures of different populations or communities?

- What resources will the program need to carry out the improvement plan?
- Has the program consulted or collaborated with key partners on potential improvement options and actions?
- In what ways could the program coordinate with other partners to achieve equity improvements that are not solely within the control or influence of the program conducting the assessment?

Additional follow-up actions help programs learn about equity impacts and whether implementation should be adjusted to realize positive outcomes. In addition, equity assessments have the potential to generate many new lessons about equity that could be helpful for other partners. Articulating plans for these actions is part of the equity assessment even though these actions occur after the formal assessment is over.

- Would sharing the equity assessment with other partners support collaboration on other policies and programs intended to benefit priority populations?
- Would sharing the equity assessment or a summary of findings with experts who were not directly involved in the assessment further promote equity through transparency and accountability?
- What measures or indicators will the program use to track progress over time? Are these disaggregated individual-level or community-level measures? Monitoring can help the program assess whether patterns or trends are in the expected direction or require course corrections.
- How and when will the organization evaluate the results of potential program changes? Evaluations focus on whether programs or policies reach their goals within a defined period. How can the organization design an equitable and inclusive evaluation?
- Who will be responsible for developing and executing monitoring and evaluation plans?
- Will the program share monitoring and evaluation results with the experts involved in the assessment or other partners? If so, how?

# 6 Data Preparation

As you're conducting your Equity Assessment in the Planning phase, we recommend documenting much of what you're finding alongside your plan for the remaining phases of the data life cycle and making that documentation as open and transparent as possible - ideally through a Data Management Plan.

A Data Management Plan describes what data will be used, how the data will be collected, processed, and managed to produce conduct your analysis, visualization or other products, and how those data products will be stored, shared, and maintained over the long-term. In other words, this Data Management Plan describes how the project team intends to address and manage each phase of the data life cycle.

Depending on the complexity of the project - your data management plan can be relatively short, and it can be used as a way to begin to engage with experts and partners that are interested in your project!

> Tip
>
> Try using the development of your Data Management Plan as a way to build relationships and trust with Tribal and community experts!
> Some ideas for how to do this include:
>
> - You can create a Technical Advisory group composed of Tribal and community experts that helps co-create the Data Management Plan with the Project Team.
> - The Project Team develops the Data Management Plan, but solicits feedback on early versions from Tribal and community experts and makes revisions according to their feedback.

A Data Management Plan should include the following sections:

1. Project Introduction & Context
2. Data Collection & Processing
3. Data Analysis & Product Development
4. Data & Product Preservation & Storage
5. Data & Product Publication & Sharing
6. Data & Product Documentation
7. Data & Product Evaluation

8. Other Potential Sections or Potential Appendices, including: Acknowledgements, Timeline, Project Roles and Responsibilities, Dataset Details, Survey Details, Future work

Since the development the Data Management Plan takes place before data are actually collected, some details, like specific analytical methods, may not be completely worked out. However, the Data Management Plan should include a clear vision and general plan for each section and include as much detail as possible.

## 6.1 Project Introduction & Context

Here you want to briefly describe the project and the mechanism(s) driving the data collection and product development. Much of this will likely be worked out during the Planning phase, including:

- What is the purpose of the selected program, policy, or process, related to this project?
- What are the objectives of the project?
- Who is the intended audience of the project?
- How do you envision the project's resulting data and products contribute to the advancement and operationalization of equity for your the program, policy, or process, related to this project?

## 6.2 Data Collection & Processing

In this section, you will identify the data you plan on collecting, how you will collect it, and how you organize, manage, and process said data once it is collected. More detailed guidance on collection and processing of data and resultant products is outlined on the Data Collection, Surveys, and Data Processing pages.

**!** Important

As you make a plan for which data you need to collect and from where - it's a great time to pause and think about what you *actually* need to answer the questions/objectives you have using an equity lens.
As a reminder - **achieving racial equity outcomes means that race can no longer be used to predict life outcomes and outcomes for all groups are improved** (Glossary)
So, as you create the list of data you want to collect for your project, it should contain:

1. Data that can represent your management question(s) or project objectives. See the Planning page for more guidance.

2. Data that can tell us something about the extent to which we are achieving equity outcomes. This may be limited to simple demographics data - but it could also be something more! Working with Tribal and community experts to decide what type(s) of data are most applicable to and reflective of their lived experiences as they relate to your management questions and project objectives is a great place to start! See the Data Collection page for more guidance.

### 6.2.1 Data Collection

A good plan will include information that is sufficient to understand the nature of the data that is collected, including:

- **Types.** A good first step is to list the various types of data that you expect to collect or create. This may include text, spreadsheets, software and algorithms, models, images and movies, audio files, and patient records.
- **Sources.** Data may come from direct human observation, laboratory and field instruments, experiments, simulations, surveys, and compilations of data from other studies.
- **Volume.** Both the total volume of data and the total number of files that are expected to be collected can affect all other data management activities.
- **Data and file formats.** Technology changes and formats that are acceptable today may soon be obsolete. Good choices include those formats that are nonproprietary, based upon open standards, and widely adopted and preferred by the larger data consuming community (e.g., Comma Separated Values [CSV] over Excel [.xls, xlsx]). Data are more accessible for the long term if they are uncompressed, unencrypted, and stored using standard character encodings.

Some questions to help guide the development of this section include:

- What data will we be collecting and/or generating?
- How and in what format will the data be collected? Is it numerical data, image data, text sequences, or modeling data?
- What file formats will be used? Do these formats conform to an open standard and/or are they proprietary?
- How much data will be generated for this project?
- Are you using data that someone else produced? If so, where is it from?
- How long will the data be collected/generated and how often will it change?
- To what extent do the data and methods of collection and use for this project abide by FAIR Principles of scientific data management and stewardship and CARE Principles for Indigenous Data Governance? If FAIR and CARE Principles are not being met - how can we modify our methods and processes data collection and use to better meet them?


Figure 6.1: FAIR Principles (Findable, Accessible, Interoperable, Reusable) within the open data movement primarily focus on characteristics of data that will facilitate increased data sharing among entities while ignoring power differentials and historical contexts. CARE Principles (Collective Benefit, Authority to Control, Responsibility, Ethics) for Indigenous Data Governance are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination. Image credit: Global Indigenous Data Allianc

### 6.2.2 Data Organization & Management

Define how and where the data will be organized and managed.

For example, your effort may require a small number of data tables and these can be effectively managed with spreadsheet programs like Excel. Larger data volumes and usage constraints may require the use of relational database management systems for linked data tables like ORACLE or mySQL, or a Geographic Information System (GIS) for geospatial data layers like ArcGIS, or computer programming languages like R or Python for large datasets that cannot be contained within standard database or GIS systems.

This section should contain just enough detail to identify basic data organization needs and plan, not the level of detail needed to build a comprehensive system or information technology project plan.

Some questions to help guide the development of this section include:

- How and where will your data be organized?
- What tools or software are required to read or view the data?
- What directory and file naming convention will be used?
- What are your local storage and backup procedures?
- Will this data require secure storage?
- Who is responsible for managing the data? Who will ensure that the data management plan is carried out?
- What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?

### 6.2.3 Data Quality & Processing

Here you will define the processes you will use to clean and prepare your data once it is collected - also known as tidying data. Tidy data are structured such that the data are easy to manipulate, model, and visualize - and getting data to the point of being tidy is often the most time consuming step of any data-intensive work. More detailed data processing/tidying steps are outlined on the Data Processing page.

Similar to the analysis step - you might not know all of the details of how the data will need to be tidied - what's important for this section is that you think through the potential methods you will need to use to make disparate datasets interoperable and tidy so that they're of acceptible quality and easy to use for your analysis and product development steps.

Some questions to help guide the development of this section include:

- Which datasets (if any) will need to be merged/combined to be made useful for your project? If this needs to be done, how will you plan on doing it?
- What are your data quality objectives/standards?
- How will you assess and establish the quality of the data you use?
- What rubric will you use to decide which data are kept and which are excluded from future steps?

> **Note**
>
> The Water Boards have a quality management system and overarching Quality Assurance Management Plan. Be sure to review this material to consider how your project falls in that that framework, and if more is needed. For example, many programs do NOT have Quality Assurance Program Plans, so this may be a step needed to occur to establish data quality objectives, etc.
> In some cases the Quality Assurance and Data Management plans can be integrated (see this USGS Quality-Assurance and Data-Management Plan as an example)

## 6.3 Data Analysis & Product Development

In this section, you will describe how you intend on using the data and the general plan or intended workflow you envision for your data analysis and/or product development phase. If you know you will use certain formulas, methods, or software for this step, you will identify them here. More detailed guidance on data analysis and product development steps are outlined on the Data Analysis and Data Visualization pages.

Some questions to help guide the development of this section include:

- What management questions are you planning on answering or informing with this data and project?
  - If operational decision making is the use please identify whiich performance measures or existing resource allocation planning processes that will be using the data (e.g., assigning inspections to staff, determining priority for compliance assurance work, etc.).
  - Please also identify any business interests that will need to be alerted to the data / products and may have concerns over its quality, etc. For example, invoicing for fees will need to have updated information, etc.
- What workflow will you use to analyse the data and/or develop the resultant data product?
- What data analysis or visualization methods or software will you use?
- What product(s) will be developed (e.g. analyses, visualizations, applications, reports, etc.)
- What opportunities will Tribal and community partners have to review and provide feedback on the data analysis or product development before it is finalized?

## 6.4 Data & Product Preservation & Storage

In this section, you will describe how and where you plan on preserving and storing data and products once they are developed. More detailed guidance on preservation and storage of data and resultant products is outlined on the Preservation & Storage page.

Some questions to help guide the development of this section include:

- How and where will you store and secure your data and resultant products (code, results, products, visualizations, applications, etc.)?
- What privacy and confidentiality issues must you address?
- What are your plans for preserving the data/products after the project is completed?
- What procedures will you use to ensure long-term archiving and preservation of your data?

- At what point will data, code/scripts, and resultant products/applications be archived or deleted?

## 6.5 Data & Product Publication & Sharing

In this section, you will describe how and where you plan on publishing, shareing and otherwise making accessibly the project's data and products once they are developed. More detailed guidance on publishing and sharing data and resultant products is outlined on the Data Sharing page.

The Water Boards typically make virtually all of the data we collect available to the public. The exceptions are confidential information (e.g., part of ongoing enforcement actions and/or formal Tribal consultations) and some personally identifiable information (PII). This section should describe any policies that will filter out data from the step of making the data publicly available and, more importantly, how the project plans to provide access to the data.

> **Note**
>
> Publishing and sharing Water Boards data and resultant products is critical for collaboration and transparency of our data, products, and workflows. Your project should be in alignment with the Water Board's Open Data Resolution: "Adopting Principles of Open Data as a Core Value and Directing Programs and Activities to Implement Strategic Actions to Improve Data Accessibility and Associated Innovation." This means:
>
> - Documenting your process throughout the project so as to make it open, transparent, and reproducible
> - Utilizing open data and open source software (e.g. Python, R) as much as possible
> - Making the data you use and code you develop transparent and accessible to the public after the project is complete, as appropriate

Some questions to help guide the development of this section include:

- What data and products will be shared, and when?
- Where and how will data and products be made open and/or accessible?
  - Datasets that are of high value should, at minimum, be published to the California Open Data Portal in the form of machine readable, well documented, maintained data.
  - Geospatial products of high value should, at minimum, be published to the California State Geoportal.
  - Code and similar products (scripts, analysis packages) should, at minimum, be published on the Water Boards GitHub in it's own, well documented, project repository.

  - For all other data or products, please indicate how the data/product will be made accessible (e.g., via search forms at SMARTS public reports page, etc.).

- Does sharing the data raise privacy, ethical, or confidentiality concerns? Do you have a plan to protect or anonymize data, if needed?
- If you collected data directly from Tribes or communities -
  - How will permission be obtained to use and disseminate the data?
  - How is informed consent being handled and how is privacy being protected?
  - How and when will you communicate what will or will not be shared?
- To what extent do the methods of publication and sharing of data, products developed through this project abide by FAIR Principles of scientific data management and stewardship and CARE Principles for Indigenous Data Governance? If FAIR and CARE Principles are not being met - how can we modify our methods and processes of publication and sharing to better meet them?

## 6.6 Data & Product Documentation

In this section, you will describe how and where every aspect of the project will be well documented. More detailed guidance on describing the project's data and products is outlined on the Documentation page.

Metadata - the details about what, where, when, why, and how the data were collected, processed, and interpreted - provide the information that enables data and files to be discovered, used, and properly cited. Metadata and other project documentation include descriptions of how data and files are named, physically structured, and stored as well as details about the experiments, analytical/visualization methods, project context, and names long-term data/product/project managers/stewards.

> **!** Important
>
> **It is generally the case that the utility and longevity of data and products relate directly to how complete and comprehensive the metadata and documentation are.**
> The amount of effort devoted to creating comprehensive metadata and documentation may vary substantially based on the complexity, types, and volume of data/products developed throughout the life of a project - but it's safe to assume (and plan for) a substantial amount of time and energy will be required to develop adequate metadata and documentation.

Some questions to help guide the development of this section include:

- What types of metadata will be produced alongside the data?

- What metadata standards will be used? Are you using metadata that is standard to your field?
- How will the metadata be managed and stored?
- What other documentation will be developed for the project and associated products (e.g. workflows, standard operating procedures, data or product use or interpretation guidance, etc.)? Where will that be stored? How will it be made accessible and shared?
- If you collected data or partnered directly from Tribes or communities -
  - Does it make sense to have these same partners review and provide feedback on your metadata and documentation materials? Doing so would help ensure that documentation is clear, simple, and accessible to a wide array of audiences.
  - When and how will you share the aforementioned documentation with your partners?

## 6.7 Data & Product Evaluation

In this section, you will describe how you will evaluate the data, products, and outcomes of the project after it is complete, to assess the extent to which the project has achieved the goals you set for it and advanced and improved equity outcomes. More detailed guidance on describing the project's data and products is outlined on the Evaluation page.

Some questions to help guide the development of this section include:

- At what point(s) during the project's life cycle will you conduct your evaluation? (You don't need to wait until the project is complete to benefit from this phase!)
- What evaluation method(s) will you use?
- How can the project design an equitable and inclusive evaluation?
- Will the project team share evaluation findings with the experts or other partners involved in the project? If so, with whom will you share it, and how?
- Would sharing the project's evaluation findings with the experts or other partners who were NOT directly involved in the project further promote equity through transparency and accountability? If so, with whom will you share it, and how?

## 6.8 Other Potential Sections

### 6.8.1 Acknowledgements

If the Data Management Plan was developed by a group that included external partners, we recommend including an acknowledgements section to acknowledge, express appreciation, and give credit to those efforts.

### 6.8.2 Project Timeline

Including a timeline for project implementation is always recommended (even though it is more of a project management tool than a data management tool) and even if specific dates are not yet known. Including a timeline helps keep oursleves accountable and makes it easier for potential partners see when their contributions, feedback, and partnership might be needed so they can plan ahead and be ready for when its their time to engage.

### 6.8.3 Project Roles and Responsibilities

If there are multiple people on the team that will be involved with project implementation, it might be a good idea to define who will be responsible for which parts of the project/data life cycle so that everyone is clear on their roles and responsibilities to this project ( even though it is more of a project management tool than a data management tool).

> Tip
>
> Spelling out project roles and responsibilities during this phase can help identify gaps and resource needs early!
> This will enable the project team, management, and/or project partners to understand the limitations and dedicate time and resources to find more team members that can help fill those gaps before the project is underway. Doing this will ultimately prevent the project from being delayed, stalled, or put on hold after time, energy, and resources have already been expended (or even wasted).

You might include a Project Roles and Responsibilities table that includes:

- Data Life Cycle Phase
- Role Title (e.g. project manager, data collection coordinator, data manager, data analysis lead, data product developer, project engagement lead, etc.)
- Name (Affiliation)
- Responsibilities (with a short list of responsibilities associated with that role)

## 6.9 Potential Appendices

### 6.9.1 Data Details

The goal of this section is making it easy for readers to see and understand the content of your data sources without having to view data directly. You might include a data schema for the datasets of interest. A data schema shows what the "guts" of your data will look like, including the identification of tables, columns/fields, data types, constraints, and relationships.

This could be provided as a single table that includes your column/field names, and data types or something much more complex that better suits the needs of your project. For a simple example, see Appendix 1 of the SWAMP Bioassessment Reporting Module Data Management Plan.

### 6.9.2 Survey Details

If your project involves collecting data through a survey, you might use this section to document your intended survey questions and possible responses or response types.

### 6.9.3 Future work

Here you might describe next steps or project ideas that are outside the scope and timelines of the current project, but that you see as being directly related to or building upon the current project.

## 6.10 Additional Resources

- DMP Tool (a web-based tool that helps you construct data management plans using templates that address specific funder requirements)
- CA Healthy Watersheds Partnership Assessment Guidance: Data Acquisition, Monitoring and Management
- USGS - Data Management Plans
- MIT Libraries - Write a Data Management Plan &
- Harvard Medical School - Data Management Plans
- University of Arizona - Data Management Plans Overview

# Part III

# Collect & Process

# 7 Data Collection

> **!** Important
>
> You should have addressed this during the Plan & Prepare phase of your process…but just in case you haven't (yet) or you need a refresher - we're restating here:
> **Achieving racial equity outcomes means that race can no longer be used to predict life outcomes and outcomes for all groups are improved** (Glossary)
> So, as you begin to collect the data for your project, be sure it includes:
>
> 1. Data that can represent your management question(s) or project objectives.
> 2. Data that can tell us something about the extent to which we are achieving equity outcomes. This may be limited to simple demographics data - but it could also be something more! Working with Tribal and community experts to decide what type(s) of data are most applicable to and reflective of their lived experiences as they relate to your management questions and project objectives is a great place to start!

The data related to your project may come from direct human observation, laboratory and field instruments, experiments, simulations, surveys, and/or compilations of data from other sources. Below we focus on data that can be downloaded from open data sources, and survey guidance.

## 7.1 Common Data Sources

Below we have provided a list of common data sources that can tell us something about the extent to which we are achieving equity outcomes.

### 7.1.1 Water Boards Data

Port over relevant water boards data and databases content?

### 7.1.2 Open Data Portals

### 7.1.3 Federal Data

### 7.1.4 Demographics Data

### 7.1.5 Common Multi-Source Data Tools

CalEnviroScreen vs. EnviroFacts vs. EnviroMapper vs. EJ Screen

## 7.2 Surveys

There may be instances where you need to collect new data using survey(s).

### 7.2.1 Survey Design

Creating surveys that yield actionable insights is all about the details - and writing effective survey questions is the first step. You do not have to be an expert to build and distribute an effective online survey, but by checking your survey against tried-and-tested benchmarks, you can help ensure you are collecting the best data possible.

Tips for Building an Effective Survey:

1. Make Sure That Every Question Is Necessary.
2. Keep it Short and Simple.
3. Ask Direct Questions.
4. Ask One Question at a Time.
5. Avoid Leading and Biased Questions.
6. Speak Your Respondent's Language.
7. Use Response Scales Whenever Possible.
8. Avoid Using Grids or Matrices for Responses.
9. Rephrase Yes/No Questions if Possible
10. Take Your Survey for a Test Drive

A good comprehensive guide for survey design can be found here: https://files.eric.ed.gov/fulltext/ED619797.pdf

https://www.qualtrics.com/blog/10-tips-for-building-effective-surveys/

### 7.2.2 Picking a Survey Software

Most Water Board staff will use Microsoft Forms which is available to all staff through the Microsoft 365 suite of applications. Microsoft Forms has a lot of advantages because of its integration with other Microsoft tools like Excel and PowerBi which allow for the survey results to be analyzed and visualized. Here is video on how to make that connection between Forms and PowerBi via Sharepoint that allows for consistent updating of results: https://youtu.be/XBFVDedwLiY?si=O161oYja-FBhG1W7

One issue with Microsoft Forms and other free software like Google Forms is that they produce wide data in Excel of Google Sheets. This type of data is more difficult to transform.

## 7.3 Data Limitations

All data have limits in what they can actually tell us, constraints on how they should be used appropriately, and biases related to initial data collection or generation - and it's important to be aware of and account for them during your project.

In most cases, the Water Boards programs were not developed or designed to collect demographic data as a matter of process which means that most will rely on a few key data sources.


Figure 7.1: Data Benefit vs Risk

Incorporating a racial equity lens during data analysis includes incorporating individual, community, political, and historical contexts of race to inform analysis, conclusions, and recommendations. Solely relying on statistical outputs will not necessarily lead to insights without

careful consideration during the analytic process, such as ensuring data quality is sufficient and determining appropriate statistical power. Disaggregation of data is also a series of tradeoffs. Without disaggregating data by subgroup, analysis can unintentionally gloss over inequity and lead to invisible experiences. On the other hand, when analysts create a subgroup, they may be shifting the focus of analysis to a specific population that is likely already over-surveilled. (Centering Racial Equity Throughout Data Integration)

Centering racial equity means paying attention to which data are highlighted and how they are framed, as well as the readability and accessibility of the communication method. This involves strategic consideration of the audience and the mode of dissemination that most effectively conveys the information. There are many ways to communicate information. These include briefs, interactive documents, websites, dashboards, social media content, data walks, posters, briefs, and infographics. Regardless of the form, content geared toward the public should avoid jargon that may be otherwise appropriate for internal program staff or academic audiences, while also using person-centered language and translating materials into languages most applicable to your community context. (Centering Racial Equity Throughout Data Integration)

Furthermore, good quality data regarding marginalized communities is often lacking, but it is still important to discuss impacts to BIPOC communities. It may be appropriate in some cases to still present or analyze this data and also present caveats for the data limitations. In other cases, it may be more appropriate to rely only on qualitative discussion based on information derived from background research and feedback from affected communities.

# 8 Data Processing

# Part IV

# Assure & Analyze

# 9 Data Analysis

# 10 Data Visualization

## 10.1 Data Visualization Tools

# 11 Demographics Data

Adding demographics data to your data project can help increase understanding of potential correlations or relationships between your data and demographic and socioeconomic characteristics of locations of interest.

Depending on what demographics data sources you decide to use, the methods needed to combine, overlay, or compare with the data you are interested in may vary. Below we outline methods of comparing demographics data to point, line, and polygon data types.

## 11.1 Demographics Data Needs Context

It's important to remember that you can always benefit from setting context before trying to communicate demographic or specific racial equity answers to questions posed by our Board, a member of the public, or a partner in this work. What your program is about, what does it do, how well does it do these things (aka Performance Report), etc. This may take a few different visualizations to help frame the context of the program mission but will help viewers understand how you are approaching the racial equity data work within the scope of your program.

Often the go-to resources for making inferences to demographic and socioeconomic characteristics is the National Census dataset and the associated American Community Survey dataset. While we are fortunate to have just updated this dataset in 2020 there are limitations and potential inaccuracies associated with relying solely on census data to enumerate demographic characteristics within a given census tract. This tool from the Department of Finance exists to measure this limitation.

A detailed example of using R programming to estimate demographics and other characteristics with U.S. census data to be used for custom spatial features is available and can be tailored to programs with the help of a data scientist proficient in R and staff familiar with the program. https://daltare.github.io/example-census-race-ethnicity-calculation/example_census_race_ethnicity_calculation.html

## 11.2 Commonly used Racial Equity Data Sources

1. CalEnviroScreen 4.0
2. United States Census Data
3. American Community Survey Data
4. Internal Administrative Data (i.e. Human Resources data)
5. Local and Regionally Collected Data

## 11.3 For a growing list of Water Boards Developed Racial Equity Data Tools:

California Water Boards Racial Equity Data Resource Hub

**More general data literacy resources on our website and in our toolkit.**

## 11.4 Collecting Data

## 11.5 Point Data

```
library(tidyverse)
library(palmerpenguins)
penguins |>                                                    ①
  mutate(                                                      ②
    bill_ratio = bill_depth_mm / bill_length_mm,
    bill_area  = bill_depth_mm * bill_length_mm
  )
```

① Take `penguins`, and then,
② add new columns for the bill ratio and bill area.

## 11.6 Lines

## 11.7 Polygons

# 12 CalEnviroScreen

CalEnviroScreen can be a helpful tool in creating visualizations and performing analysis as it provides a number of index, as well as a "rolled-up" score that combines environmental and demographic data together. However, there can be things to consider, a couple of which are discussed below.

## 12.1 Missing Values for CalEnviroScreen Scores

Users conducting an analysis with the CalEnviroScreen (CES) 4.0 dataset should be aware that it contains missing values, both for individual indicators and overall CES scores. These missing values are distinct from zeros, which are also in the CES dataset. For more information about the missing (and zero) values, see the data dictionary (calenviroscreen40resultsdatadictionary_F_2021.pdf) that accompanies the CalEnviroScreen 4.0 results Excel workbook, available for download as a zip file here.

In the CES 4.0 data (for the version available as of April 2023), the shapefile containing CES 4.0 scores (available here) encodes these missing values as negative numbers (-999 for most variables, and -1998 for one variable). The Excel workbook containing CES 4.0 scores (available here) encodes these missing values as NA. Users should account for these missing values – and their different encodings – as needed when doing any analysis using CES data. Also, note that the CalEnviroScreen 3 shapefile (June 2018 update version) encoded missing values as 0, so users should be aware of this change if/when updating an analysis from CES 3 to CES 4.0 data.

## 12.2 Inconsistent Census Tract Boundaries in CalEnviroScreen 4.0 Shapefile

In the CES 4.0 data (for the version available as of April 2023), the shapefile containing CES 4.0 scores (available here) uses a simplified version of the polygons that represent 2010 census tracts. The boundaries of the census tracts defined by these simplified polygons do not always align with the boundaries of neighboring census tracts, resulting in slight gaps or overlaps between some neighboring census tracts. These inconsistencies are not likely to have a significant impact on most uses of the CES data, but they could impact some types of

analysis based on CES data. For example, when assessing sites or facilities based on the CES score of the census tract they are located in, sites located near a census tract boundary could be associated with more than one census tract (and more than one CES score) in areas where there are overlapping census tract polygons, or not associated with any census tract (and no CES score) in areas where there are gaps between census tract polygons. This issue may be addressed in a future release of the CES dataset; in the meantime, a possible workaround is to use the official 2010 census tract boundaries from the US Census Bureau for any calculations, then use census tract IDs to tie this information to the associated CES score for each tract.

# 13 Turning Data into Information

Turning Data into Information is Challenging - Even More so with Racial Equity Data Work

Turning data into information in the context of racial equity involved navigating complex ethical considerations. The process requires an understanding of the potential impact on Black, Indigenous, and other People of Color (BIPOC) communities and the responsibility to mitigate perpetuating or reinforcing biases. Upholding ethical standards requires a commitment to maintaining privacy, accessibility, and fostering transparency throughout the data transformation process. Additionally, acknowledging the limitations of the data and being transparent about potential biases is essential for maintaining the integrity of the data and information generated and shared. The transformation of racial equity data into meaningful information requires a thoughtful and intentional approach which we will highlight in the next sections.

For example, many programs will rely on demographic and socioeconomic data, like those collected from the U.S. Census and the American Community Survey (ACS). Because the ACS is based on a sample, rather than all housing units and people, ACS estimates have a degree of uncertainty associated with them, called sampling error. In general, the larger the sample, the smaller the level of sampling error. To help users understand the impact of sampling error on data reliability, the Census Bureau provides a "margin of error" for each published ACS estimate. The margin of error, combined with the ACS estimate, gives users a range of values within which the actual "real-world" value is likely to fall. It is important to acknowledge this uncertainty up front to be transparent with your audience about the data and conclusions you are drawing. For more information on using ACS data please see the Understanding and Using American Community Survey Data: What All Data Users Need to Know Handbook. Also for an interactive tool that measures the potential inaccuracies associated with relying on census data to enumerate demographic and socioeconomic characteristics in California please explore: https://cacensus.maps.arcgis.com/apps/webappviewer/index.html?id=48be59de0ba94a3dacff1c9116df8b37.

Another example of how to tell your data story in a transparent way can be found on the Office of Environmental Health and Hazard Assessment CalEnviroScreen 4.0 Race and Equity Analysis. https://storymaps.arcgis.com/stories/f555670d30a942e4b46b18293e2795a7

# Part V

# Preserve & Store

# 14 Preservation & Storage

# Part VI

# Publish & Share

# 15 Data Sharing

add guidance on open/sharing practices & locations - GitHub, Open Data Portals, etc.

# Part VII

# Discover & Integrate

# 16 Evaluation

After a project is complete (or has gone through a complete iteration) - it's important to assess the project and evaluate the extent to which it has achieved the goals you set for it and advanced and improved equity outcomes. There are several ways to achieve this. Below is a list of methods one might consider utilizing for this phase of the data life cycle.

Depending on what you discover during the evaluation process, you may need to undergo a second phase or iteration of the project to integrate lessons and get closer to achieving the original goals and desired equity outcomes.

> Tip
>
> You don't need to wait until the project is complete to benefit from using these tools! It may be helpful to use these tools at multiple points during your project or process.
> ***When in doubt - make time to use these tools early and often!***

## 16.1 Root Cause Analysis

## 16.2 Results-Based Accountability (RBA)

## 16.3 Performance Accountability

## 16.4 Other Racial Equity Lens Evaluation Questions

# Part VIII

# Describe

# 17 Documentation

Add guidance on documentation, reproducibility & transparency, oh my!

This is intended to be a living document and providing documentation is extremely important.

> **!** Important
>
> Remember - if your project or process is not well-documented to the point of being largely reproducible - it's incomplete!

# Part IX

# Use Cases

# 18 SWAMP

change page title to short description of use case

# 19 Demographics

# 20 Resources

Here you will find a curated list of presentations, webpages and other resources related to the development, implementation and scaling of the principles and practices outlined in this Data Equity Handbook.

All Water Boards authors are **bolded** below.

## 20.1 Other Equity & Data Handbooks / Toolkits / Guides

- Water Boards Developed Racial Equity Data Tools
- Water Boards Environmental Justice Roundtable Resource Catalog
- GARE Racial Equity Toolkit: An Opportunity to Operationalize Equity
- Beyond Compliance Network Advocacy Toolkit
- Academic Data Science Alliance Data Science Ethos
- CA Water Boards College of Water Informatics Data Toolkit
- 500 Women Scientists Guide for Inclusive Scientific Meetings

## 20.2 Presentations

Analyzing Water Boards and Demographic Data for Equity. Jun 2024. Hannah Cushman Garland. State Water Board Racial Equity Data Subcommittee Webinar. Recording | Download and Use the Code | View Code

## 20.3 Relevant Government Documents

### 20.3.1 California Water Boards

- Racial Equity Resolution (2021) - English, Español
- Racial Equity Action Plan (2023) - English, Español
- Racial Equity Resolution Annotated References (2021)

### 20.3.2 California Office of Data and Innovation

- California Data Strategy Report (2024)
- California Data Strategy (2020)

### 20.3.3 The White House

- The Environmental Justice Science, Data, and Research Plan (2024)
- Year of Open Science Fact Sheet (2023)

## 20.4 Relevant Literature

O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., Vera, L., Jennings, L.L., Downs, R.R., Antognoli, E., ten Brink, T., Halmai, N.B., David-Chavez, D., Carroll, S.R., Hudson, M. and Buttigieg, P.L. (2024) 'Earth Science Data Repositories: Implementing the CARE Principles', *Data Science Journal*, 23(1), p. 37.

## 20.5 Websites

- Racial Equity at the Water Boards
- Openscapes at the Water Boards
- California Office of Data and Innovation
- Government Alliance on Race and Equity (GARE)

# 21 Inspiration

The impetus for developing this Data Equity Handbook began in August 2020 when the State Water Resources Control Board (State Water Board) publicly acknowledged that the historical effects of institutional racism must be confronted throughout government, and it directed staff to develop a priority plan of action.

Since then, the State Water Board, it's Office of Information Management and Analysis (OIMA) and OIMA's many internal and external partners have been developing and compiling material, and adding to this Data Equity Handbook as time and bandwidth allow.

This Data Equity Handbook is inspired by many sources, including:

- Water Boards Racial Equity - Resolution and Related Actions
- Government Alliance on Race and Equity (GARE)
- Openscapes and their Approach Guide
- NOAA Fisheries (NMFS) Open Science Resource Book