Data Disaggregation:
A Blueprint for Government Agencies

To tackle the many inequities in our society, policymakers, advocates, and the public need data that are complete, accurate, and disaggregated. Disaggregation means collecting and reporting data on subgroups, such as by race, ethnicity, gender, and age, so the data accurately reflect reality for different subgroups of people.

When data are not disaggregated, broad trends can hide inequities. For instance, in education, Southeast Asian American students face significant barriers to college access and success. But if education data are not disaggregated, this reality for Southeast Asian American students is masked by relatively higher average rates of college success by the broader group of Asian American, Native Hawaiian, and Pacific Islander (AANHPI) students.

The need for more disaggregated data is a theme of The Leadership Conference Education Fund's report "Data for Equity: A Review of Federal Agency Equity Action Plans," as well as many of The Education Fund's previous reports on civil rights data collection — and it has been a priority for civil rights advocates for decades. The Biden administration's Equitable Data Working Group's final report also reinforced the principle that disaggregated data should be the "norm" while protecting privacy.

The Data for Equity report, which was based on a review of dozens of Agency Equity Action Plans ("Action Plans") released in April 2022, makes a number of specific recommendations about how agencies can improve data disaggregation:

Update OMB standards on race and ethnicity

The biggest reform needed to ensure meaningful data disaggregation across the federal government is updating OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity ("standards"). These standards, which have not been updated since 1997, guide how all federal agencies collect information about race and ethnicity.

The racial and ethnic categories in the standards fail to reflect the identities of many people, including people of Middle Eastern and North African heritage; subgroups of Asian Americans and Native Hawaiian and Pacific Islanders; and people of Hispanic or Latino heritage.

Communities and advocates have long urged OMB to revise its standards to reflect the identities of these and other groups. Unfortunately, although OMB has twice solicited public comments on the standards and announced that a final decision would come by December 2017, it has not yet updated them.

On January 27, 2023, OMB published a [Federal Register Notice](#) of its initial proposals for updating the standards. OMB should strive to complete this vital update by the announced date of summer 2024.

Agencies can and should collect more disaggregated data on race, ethnicity, gender, and other identities even without changes to the OMB standards

While agencies cannot change the minimum reporting categories they use to collect racial and ethnic data before OMB changes its standards, they can and should collect more disaggregated data about race and ethnicity within the current standards, as well as disaggregated data about gender and other identity categories. The chief statistician of the United States has published [guidance](#) on best practices and disaggregation under the current standards.

Several agencies' Action Plans discuss their plans to increase disaggregation — or reveal the need to do so.

Equal Employment Opportunity Commission (EEOC) plans to increase disaggregation of race, ethnicity, and gender data

The [EEOC's Action Plan](#) notes that more disaggregated data could help it better understand persistent forms of discrimination — for instance, hiring discrimination against Black workers. The agency's Action Plan highlights its plans to update demographic categories on internal EEOC forms, specifically regarding gender self-identification; perform

deeper analysis of existing data; and determine whether to add additional demographic categories to EEO data collections.

U.S. Department of Health and Human Services (HHS) must require Affordable Care Act-covered entities to collect disaggregated demographic data

[HHS's Action Plan](#) refers briefly to the fact that the agency lacks the data necessary to "consistently identify and address inequities in health and human services." However, the Action Plan does not discuss that absence in detail or say how the agency will rectify it.

One way it should do so is by requiring entities covered by the Affordable Care Act (ACA) to collect disaggregated data. In August 2022, the agency released a [Notice of Proposed Rulemaking](#) (NPRM) on Section 1557 of the ACA, which prohibits discrimination on the basis of race, color, national origin, sex, age, or disability in health programs. That NPRM does not require entities covered by the ACA to collect disaggregated demographic data. As The Leadership Conference and 100 organizations noted in [comments](#) to HHS on the NPRM, the final rule must require the collection of disaggregated data.

U.S. Department of Housing and Urban Development (HUD) needs to collect more disaggregated data in several areas

[HUD's Action Plan](#) notes that it is "likely" that LGBTQ+ households of color face compounding, intersectional discrimination in the realm of homeownership and associated wealth accrual — but the agency does not currently have the data to determine if this is

true. In addition to making it impossible to assess the current baseline, this lack of disaggregated data will make it more difficult for HUD to implement its plan to tackle housing discrimination through increased culturally responsive and linguistically competent housing counseling, among other approaches. HUD must collect or use existing data on home purchasing to understand and address discrimination in that field.

HUD also identifies a goal of improving its data on the race, ethnicity, and gender identity of people experiencing homelessness and to gather the data in a culturally sensitive and trauma-informed manner. The agency intends to do this by updating the Homeless Management Information System data collection requirements for race, ethnicity, and sexual orientation and gender identity (SOGI).

Social Security Administration (SSA) should expand beyond binary M/F gender designations

Currently, SSA only allows "M" or "F" gender designations; there is no indicator for nonbinary people. If a person's underlying identity document provides a nonbinary marker like "X," SSA cannot process the application automatically; the applicant has to go to a field office and must still identify themselves as either male or female to satisfy SSA's computer systems. Changing from one gender to another is burdensome. SSA's Action Plan states that the agency intends to "explore self-attestation for changing sex designation" and training staff on unconscious bias, but it does not propose actually changing those policies it identifies as barriers. It should do so.

National Science and Technology Council (NSTC) Subcommittee on Equitable Data and the Census Bureau are exploring oversampling American Indian and Native Alaskan people

According to the Council of Economic Advisors' Action Plan, through its work on the Equitable Data Working Group (now known as the NSTC Subcommittee on Equitable Data) it is working with the Office of Management and Budget to explore the possibility of oversampling people from American Indian/Native Alaskan groups in the Current Population Survey and possibly other Census Bureau surveys. Oversampling can be necessary to have sufficient data to analyze a phenomenon in smaller communities, but it is important that it be done without overburdening the communities in question. The NSTC Subcommittee on Equitable Data, the bureau, and the other agencies should complete their exploration and make changes as needed.

Office of National Drug Control Policy (ONDCP) plans to improve data on substance use

ONDCP's Action Plan identifies the lack of robust disaggregated data on substance use as an "overriding challenge." For instance, until recently, it says, data were not collected on SOGI categories, and the "Hispanic" category includes many subpopulations that include widely different patterns of drug use. Surveillance data on drug use also tends to undercount people experiencing homelessness, sex workers, incarcerated people, and residents of rural areas.

One of the agency's goals is to improve data systems and promote disaggregation to better understand which subpopulations are at high risk for drug use. For instance, the agency will work with OMB to add more demographic data collection to progress report forms for grant recipients.

Department of Labor (DOL) plans to improve disaggregation of data on unemployment insurance and apprenticeships

One of DOL's Action Plan goals is to improve equity in the unemployment insurance (UI) system. A lack of disaggregated data currently prevents regular equity tracking. DOL's Action Plan notes that it appears people of color and people who are low-income, limited English proficient, and have lower levels of education are less likely to receive UI benefits, although "different data sources sometimes point in different directions."

DOL's plan includes creating new indicators of UI program access using application, denial, and timeliness rates broken out by demographic characteristics and recruiting more states to share their data (currently only five do) so it can produce disaggregated federal indicators of access to UI benefits for the first time.

DOL's Employment and Training Administration also plans to expand the scope and disaggregation of data it collects on apprenticeships and other programs.

The Leadership Conference
Education Fund