II. BACKGROUND

In the earliest months of the COVID-19 pandemic, the Center Director, Ibram X. Kendi, feared that the virus was disproportionately harming and killing people of color.[22] He could not get data to confirm or deny this fear, however, as nationwide racial and ethnic data on COVID-19 outcomes were practically nonexistent.[23] At the time, some public officials referred to COVID-19 as the "great equalizer."[24] In reality, public health emergencies disproportionately devastate under-resourced communities of color, and COVID-19 was no exception.[25] But the lack of reliable data regarding COVID-19 outcomes by race and ethnicity was a major barrier to creating equitable policy interventions.[26] Dr. Kendi began publishing pieces in *The Atlantic* urgently calling for COVID-19 race and ethnicity data.[27] As the months wore on and people continued to die and suffer, various states slowly began to release some racial and ethnic data for COVID-19 outcomes in an uncoordinated, piecemeal manner, but nationwide trends remained unknown, and there were many gaps in the data.

With limited information forthcoming from any national agency, Dr. Kendi partnered with *The Atlantic's* COVID Tracking Project in the early Spring of 2020 to quickly assemble a team of journalists, researchers, scholars, and volunteers to fill this gap, and the COVID Racial Data Tracker (CRDT), the Center's first research project, was born. From April 2020 to March 2021, the CRDT team painstakingly worked to manually collect the available racial and ethnic COVID-19 data reported by state and U.S. territory.[28] The CRDT, which constituted the most complete and timely source of COVID-19 race and ethnicity data at the time, confirmed that COVID-19 reflected the long-standing racial and ethnic health inequities that plague the U.S. population: Black, Brown, and Indigenous people were disproportionately suffering and dying from the virus.

Nevertheless, the CRDT data could only be as complete and accurate as the data the team was collecting from each state. The CRDT team encountered a number of data quality challenges due to the varied ways states reported (or failed to report) data, and responded to those challenges as best it could. While the CRDT team's work successfully demonstrated the existence of racial and ethnic inequities in COVID-19 outcomes, state data quality issues prevented the team from revealing the full extent of the inequities.[29] The CRDT is therefore a useful case study of how state data sources are an insufficient substitute for single, standardized, nationwide data, and how difficulties in combining data reports from the different states and territories, each with their own reporting practices, can introduce errors and obscure evidence of racism. This case study allows us to identify better practices for collecting and reporting race and ethnicity data and the measures needed to fill existing gaps in the data.

Aware that the data quality issues the CRDT team faced were not unique to COVID-19 or the public health context, Center staff then decided to expand the CRDT's work by establishing the Racial Data Tracker (RDT) to continue this investigation. The Racial Data Tracker team is advancing Racial Data Science, a multidisciplinary field marrying antiracist research and data science that involves the application of mathematics, statistics, computer science, visualizations, storytelling, and social science methods to large volumes of data. To do this, the RDT team is working to amass the largest online, publicly accessible collection of racial and ethnic data at the local, regional, and national levels, spanning key issue areas including health, education, employment, politics, housing, and the criminal legal system. The RDT team began by collecting racial and ethnic data on houselessness, criminal arrests, and police violence.

The RDT team's experience sheds more light on many of the data quality issues identified by the CRDT team. The RDT team tried to gather data from numerous existing national, state, and local datasets in order to obtain the clearest picture of racial and ethnic inequities across important issue areas. The RDT team's experience confirms that racial and ethnic data collection and reporting in the United States are frequently left to state, local, and nonprofit entities, whose participation is often voluntary, inconsistent, incomplete, and poorly enforced.

The incompleteness of racial and ethnic data is a major barrier to crafting equitable policies and meaningfully measuring progress toward racial equity.[30] "If we can't see racial disparities, then we can't see the racist policies behind any disparities and deaths."[31] And if we can't identify racist policies—that is, policies that produce or sustain racial inequity—we cannot dismantle them.[32] The CRDT and RDT teams' experiences underscore the urgent need for a single, standardized, nationwide system of collecting and reporting data by race and ethnicity. Such a system would

allow policymakers to see whether and how particular policies and practices are racist, and respond with informed solutions, paving the way for eradicating "the original American virus: racism."[33]

The following sections of this Report examine the data collection processes of the CRDT and RDT teams from the Spring of 2020 to the Summer of 2021. Their experiences illustrate the deficiencies of existing publicly available racial and ethnic data[34] for COVID-19, houselessness, criminal arrests, and police violence. We describe the current challenges of collecting robust and accurate racial and ethnic data, demonstrate ways in which incomplete data obscure racial inequities and erase evidence of racism, and offer antiracist policy recommendations.