Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

EPISTEMOLOGIES OF MISSING DATA: COVID DATA BUILDERS AND THE PRODUCTION AND MAINTENANCE OF MARGINALIZED COVID DATASETS

Youngrim Kim
Rutgers University

Megan Finn
American University

Introduction

During COVID-19, countless dashboards have served as central media where people learn critical information about the pandemic. Varied actors, including news organizations, government agencies, universities, and NGOs created and maintained these dashboards, conducting the onerous labor of collecting, categorizing, and taking care of COVID data. This study uncovers different forms of data practices and labor behind the building of these dashboards, based on in-depth interviews with volunteers and practitioners across India and the United States who have participated in COVID dashboard projects.

Specifically, we are interested in projects that have focused on underrepresented or missing COVID data such as COVID cases in prisons and long-term care facilities, racial/ethnic breakdown of cases, as well as deaths due to COVID enforcement. These data builders employed sometimes creative, sometimes mundane and laborious data practices to not simply collect, but to produce these data that are often absent or obscured in the official COVID dataset. In this process of data production, dashboard builders problematized the state's hegemonic ways of quantifying death and illness, as well as grappling with the questions of how certain data is collected, who/what is missing from the dataset, and how these data voids shape and manipulate our understanding of the pandemic. Through in-depth interviews with 74 data builders who participated in COVID dashboard projects, this paper demonstrates the range of underrepresented and messy COVID data that these data builders identified and engaged with – disappearing data, obscuring data, and disregarded data that these builders repaired and maintained to render them useful. Such critical engagement with

Suggested Citation (APA): Kim, Y., Finn, M. (2023, October). *Epistemologies Of Missing Data: Covid Data Builders And The Production And Maintenance Of Marginalized Covid Datasets*. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from http://spir.aoir.org.

missing COVID data reveals different data injustices that have tremendous potential to affect future pandemic preparation and intervention efforts.

Missing Data and the Politics of Counting in COVID-19

Under- or misrepresented datasets have been extensively investigated by rich feminist and indigenous scholarship that has paid attention to systematic biases and institutional neglect that render certain populations, activities, and other aspects of human lives absent from state records (Wernimont, 2018; Tuana, 2007; Carlson, 2021). These works emphasize how the absence or lack of certain data is both productive and systematic, revealing the value systems of our society: whose lives are deemed worthier than others, what is considered more countable and profitable, and thus better accounted for in the state's quantification practices?

Contributing to this feminist scholarship, D'Ignazio (2023) provides a critical definition of "missing data": "those that are neglected to be prioritized, collected, maintained, and published by institutions, despite political demands that such data should be collected and made available." Hence, identifying and collecting such "missing data" is a central form of data activism that demands critical reflection on why certain knowledge is absent and/or overlooked. As a result, D'Ignazio emphasizes that attention to missing data is an active manifestation of state ignorance as well as a call for a structural repair.

The COVID-19 pandemic is a case where we have witnessed both an overabundance and a dearth of critical COVID data. Milan and Treré (2020, 2021) pointedly speak about the COVID data divide – a term describing Global North as having the privilege of being "data-rich" while marginalized communities suffer from "data poverty." Pelizza et al. (2021) also spoke about how undocumented migrants have been actively neglected by the state's COVID counting, making them inaccessible to basic healthcare and civil rights protections. Problematizing such hegemonic quantification of COVID-19, disaster studies scholars emphasized the need for more inclusive and humane data practices that account for care, rather than control and governance (Soden et al., 2022). Mattern (2022) in particular, speaks about how some common data practices such as dashboards and mapping could have the potential to show the political meaning of missing data by telling stories that the official data sources neglect. The COVID dashboard projects that we examined actively attend to these data voids in a way that Mattern describes, as shown in the findings below.

Method

We conducted semi-structured, in-depth interviews from March to November 2021 with 74 participants who were involved in dashboard projects in the United States and India. We recruited interviewees from 11 different dashboard projects, including The COVID Tracking Project, Johns Hopkins University's Coronavirus Resource Center, and The Marshall Project's Prison COVID data. This paper specifically focuses on dashboard projects that collected COVID data in specialized contexts: COVID racial data, COVID in long-term care facilities, prison COVID data, racial violence during COVID, and deaths due to COVID enforcement. Each interview took between one to two hours and was conducted over Zoom.

The Landscape of Missing COVID Data

One type of missing data we learned from our interviews was disappearing and ephemeral data. COVID cases in vulnerable communities such as prison systems and long-term care facilities were often removed or reverted without clear explanations. For instance, reporters at The Marshall Project, who tracked COVID numbers in US prison systems, noticed in March 2021 that the cumulative numbers of COVID cases reported by the Federal Bureau of Prisons (FBP) were in fact, decreasing. When they investigated this issue, reporters discovered that the FBP was removing people who had been released from prisons from their totals: “So we still don’t have a great sense of the total number of people who were ever infected in federal prisons,” said our interviewee. To capture such disappearing and ephemeral data, it was crucial for data builders to meticulously track and probe any changes in state prisons’ COVID pages. At times, this involved building their own web tracking software that would monitor a certain part of a webpage every second, as well as manually contacting public information officers of each state prison to question the reasoning behind suspicious changes in numbers.

The second type of missing data that we identified is obscuring data. Here, we argue that the concept of “missing data” does not only indicate an absence of certain data, but also those that are made meaningless to accurately reflect the extent of the pandemic. One case that shows such an extended understanding of missing data is lumped datasets from long-term care facilities. The COVID Tracking Project’s long-term care team identified that the states’ reporting of COVID cases in these facilities often report only the aggregate number of cases of different nursing facilities, without breaking down by type such as nursing homes, assisted living, family-type homes, or adult homes. Knowing that assisted-living facilities have fewer resources and are thus more vulnerable to COVID than nursing homes, only showing the combined number of cases obscures us from seeing where inequality is happening. To tease out these numbers, the long-term care tracking team compared multiple state datasets with reports from the Centers for Medicaid and Medicare to produce facility-type breakdowns.

Lastly, some dashboard builders invented new categories that are disregarded from official COVID datasets to highlight COVID-induced injustices, particularly in extremely impoverished communities. For instance, the Non-virus Death Database, built by academics and public interest technologists in India, tracked the number of deaths due to COVID enforcement such as suicides, domestic violence, and starvation due to lockdowns. Here, counting missing data became a means to studying the conditions of these deaths, and a process to tell poignant stories of the underserved communities’ pandemic experience.

References

Carlson, B. (2021). Data silence in the settler archive: Indigenous femicide, deathscapes and social media. In S. Perera & J. Pugliese (Eds.), *Mapping*

Deathscapes: Digital Geographies of Racial and Border Violence (pp. 84-105). Routledge.

D' Ignazio, C. (2023). *Counting Feminicide: Data Feminism in Action*. MIT Press.

Mattern, S. (2021). How to map nothing. *Places Journal*. Available at: https://placesjournal.org/article/how-to-map-nothing/

Milan, S. & Treré, E. (2020). The rise of the data poor: The COVID-19 pandemic seen from the margins. *Social Media + Society, 6*(3): 1-5. https://doi.org/10.1177/2056305120948233

Milan, S. & Treré, E. (2021). A widening data divide: COVID-19 and the Global South. In S. Milan, E. Treré, & S. Masiero (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies, and Resistance in the Datafied Society* (pp. 24-28). Institute of Network Cultures, Amsterdam.

Ọnụọha, M. (2016). The library of missing datasets. Available at: https://mimionuoha.com/the-library-of-missing-datasets.

Pelizza, A., Milan, S., & Lausberg, Y. (2021). The dilemma of undocumented migrants invisible to COVID-19 counting. In S. Milan, E. Treré, & S. Masiero (Eds.), *COVID-19 from the Margins: Pandemic Invisibilities, Policies, and Resistance in the Datafied Society* (pp. 70-78). Institute of Network Cultures, Amsterdam.

Soden, R., Wernimont, J., & Knowles, S. G. (February 24, 2022). Accounting for care in times of crisis. *Items*. Available at: https://items.ssrc.org/covid-19-and-the-social-sciences/disaster-studies/accounting-for-care-in-times-of-crisis/

Tuana, N. (2006). The speculum of ignorance: The women's health movement and epistemologies of ignorance. *Hypatia, 21*(3): 1-19. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1527-2001.2006.tb01110.x

Wernimont, J. (2018). *Numbered Lives: Life and Death in Quantum Media*. MIT Press.