

# COVID-19 genomic data sharing, diverse Enough?

Nathanael Sheehan

2023-01-16

The question of “how diverse is diverse enough?” is a fundamental concern in the philosophy of science, particularly in light of the recent rise of open science (OS). The practical argument for epistemic diversity suggests that a diverse group of scientists, with different backgrounds, experiences, and perspectives, will lead to a more comprehensive and accurate understanding of the phenomena being studied. The normative argument for epistemic diversity, on the other hand, posits that diversity is necessary for the fair and just production of knowledge. This argument holds that certain groups, such as marginalised communities, have been historically excluded from the scientific process and that increasing diversity from these voices within the scientific community can ensure that the knowledge produced is more inclusive and representative of the pluralist perspectives of scientists as a whole. However, the extent to which diversity is truly achieved and its limits are still subject of ongoing philosophical inquiry and debate.

The globally coordinated Covid-19 genomic surveillance infrastructure serves as a glaring example of the limitations and failures of epistemic diversity in scientific practice. Despite the rhetoric of global collaboration and coordination in the name of finding solutions for the pandemic, the reality is that genomic science is deeply rooted in the histories and ongoing practices of European and North American imperialism. These imperialistic tendencies manifest in forms of socio-scientific paternalism, the exploitation and invisibilisation of marginalised labor, and the ongoing dissonance between colonised nations and their colonising counterparts. These issues not only impede the localisation and cultural relevance of genomic science, but also severely impact the translation of data into knowledge and the deployment of technology. GISAID and the Covid-19 Data Portal, while both claiming to promote open access to genetic data on the virus, differ significantly in their policies and approaches to data sharing, with GISAID requiring authentication and permission for data sharing, and the Covid-19 Data Portal having no such restrictions. The examination of these infrastructures reveals the complex relationship between governmentality, values, incentives, and epistemic diversity in OS initiatives.

In this paper, we propose a mixed methods approach to investigating the relationship between epistemic diversity and the governance of the Covid-19 genomic surveillance infrastructure. Specifically, we will employ a quantitative analysis of the two infrastructures in terms of their support for epistemic diversity, operationalised as the number of authors, institutions, ontologies, and geographies represented in global publications that make use of data from either. Additionally, we conduct semi-structured interviews with members and users of each database to gain a more nuanced understanding of the ways in which these infrastructures are promoting or hindering epistemic diversity. Through this approach, we aim to provide a more situated and localised analysis of the complex relationship between governmentality and epistemic diversity, and to consider the role of institutions and their policies in shaping the diversity of scientific knowledge.