Challenges of Creating Open Datasets about COVID-19 Policies in Brazil

by **Andreza Aruska de Souza Santos** | Oxford School of Global and Area Studies, University of Oxford | andreza.desouzasantos@lac.ox.ac.uk

and **Anna Petherick** | Blavatnik School of Government, University of Oxford | anna.petherick@bsg.ox.ac.uk

In December 2019, a new respiratory virus was reported in the province of Wuhan, China. Since then, at least 71,351,695 million people have been infected worldwide and 1,612,372 people have died, as of 15 December 15, 2020 (https://covid19.who.int). In the intervening year, as the virus has spread, health systems have struggled, and as democratic and autocratic governments alike have restricted individual freedoms, lives have changed across the globe. Research has also changed. Scientists have found new ways to collaborate, to curate and share datasets, and to publish research on the pandemic at an unprecedented speed. This article addresses some behind-the-scenes challenges of collaborative work across borders and disciplines, from the standpoint of two female, early-career researchers. The perils and privileges of the past few months are summarized within three themes: (1) success and complications of scientific collaborations, (2) pressures relating to the academic job market pre- and post-COVID-19, and (3) politicization and communication of research findings, focused on the Brazilian context.

New Methods and Collaborations across Borders and Disciplines

From different departments at the University of Oxford, the authors have worked independently, curating complementary datasets of Brazilian public policies. The decentralized policy response to COVID-19 in Brazil, where each of the country's 5,568 municipalities and 27 state governments can pass their own decrees to enact or ease preventative measures, has meant that researchers seeking an overview of Brazil's nonpharmaceutical interventions (NPIs) have faced a dizzying task. Each jurisdiction has typically passed dozens of decrees. Meanwhile, the policies of geographically superimposed units of governments have frequently been contradictory. To reduce the challenge of studying Brazil's NPIs for the research community, we have both set about collecting and synthesizing policy data. Andreza and the CADDE Centre (Brazil-UK Centre for [Arbo]virus Discovery, Diagnosis, Genomics and Epidemiology) initiated a collaboration with a Brazilian Municipal Association (Confederação Nacional de Municípios) to collect survey responses from Brazilian mayors, especially those leading small and medium-size towns (a focus that corresponds to 94 percent of all Brazilian cities). In the survey they collected NPIs and the dates they were enacted for over 4,027 Brazilian cities, uniquely adding to the complexity of data collection in Brazil.[1] Meanwhile, Anna and other members of the Oxford COVID-19 Government Response Tracker (OxCGRT) project assembled a team of 77 Brazilian volunteers to systematically code the strength of policies using ordinal scales, in (at the time of writing) 15 policy domains. This dataset is updated in real time and freely available on GitHub.[2] Every data point passes through a reviewing system within a fortnight. The data is currently available for 81 jurisdictions, covering the

---

1   Andreza Aruska de Souza Santos et al., "SARS-CoV-2 non-pharmaceutical interventions in Brazilian municipalities" (dataset), Dryad, October 22, 2020, https://doi.org/10.5061/dryad.vdncjsxs2.

2   A. Petherick, B. Kira, L. Barberia, T. Boby, R. Goldszmidt, and M. Luciano, "Brazil's COVID-19 Government Response Policies," Blavatnik School of Government, https://www.bsg.ox.ac.uk/research/research-projects/brazils-covid-19-policy-response; dataset, https://github.com/OxCGRT/Brazil-covid-policy/tree/master/data.

three levels of the federal system across the country: the federal government, all state governments, and the capitals and second cities of each state.

Curating datasets presented us with challenges as well as gains. Working in teams that straddle the Atlantic has meant optimizing time and allowed for an around-the-clock work rhythm. This rapid production of information has been and continues to be necessary for the research and policy community as the pandemic advances, and Brazil remains one of the most affected countries. On a personal level, the large collaborations we describe were, while exhausting, no doubt also greatly beneficial. In a moment when we found ourselves in social isolation and with a sense of worry (in the case of Andreza her family and friends are mostly living in Brazil, with many having tested positive for COVID-19), the research projects in which we have been engaged have given us sense of agency during crisis. We are also indebted to our collaborators for the skills they have imparted, and for their enthusiasm and commitment.

At the same time, we have found that features of the academic ecosystem have made it occasionally difficult to move at the pace that a pandemic response requires. Getting funding and ethics approval is a time-consuming effort, and at the beginning, we set projects up before COVID research grants even existed. Though ethics committees and funding bodies have prioritized COVID-19 related research, the pace of pandemic spread, and the rapidly adjusting dynamics of pressing research and policy questions, have greatly surpassed that of academic administration. Sending money abroad quickly and getting questionnaires and phone surveys into the field, in the context of lockdown, have been hard. Delays at the start have translated into even more intense working hours to catch up. As two early-career researchers who have pushed aside pre-pandemic research projects and publication streams, we have been constantly aware that ploughing time and energy into data collection and curation does not always contribute to career progress.

Gains That Do Not Translate in the Current Academic Landswcape

The greater gain of data curation for the community of scientists and health professionals is indisputable, and to be sure, there is a growing trend in academia to acknowledge the value of datasets on their own, such that they do not necessarily need to be connected to a paper. However, journal publication remains as the only medium with an impact factor, a key signal of achievement and contribution in academic careers. While both of us have managed to eke out slivers of time to publish, we certainly would have generated more academic publications as data users than as data producers. Andreza has collaborated with epidemiologists and virologists and coauthored several high-impact publications that use the data she has been curating. Anna and her collaborators in Brazil have run panel surveys in nine state capitals, to ascertain opinions and adherence to the policies, which the OxCGRT project codes, and published working papers to inform policymakers. Yet we both harbor a slightly forlorn sense of missed research opportunities, of paths left unpursued because we lacked time to dig into our own datasets.

The fact matters that data generation and sharing do not tick widely recognized boxes in a CV for the academic marketplace, especially for early-career researchers.[3] And yet many of those with the skills and who are able to find the time to embark on such projects are indeed young scholars. When schools and nurseries closed, senior colleagues, often with caring responsibilities, tend to be less able to dedicate extra hours to long meetings, to around-the-clock data collection and interpretation, and to daily engagement with and explanation of the datasets to users. Early career researchers, either because of the privilege of a life that still affords some spare time, or because the precariousness of their positions encourages taking on extra tasks, have contributed a great deal of pandemic-related research and associated activities.

[3] Adam J. Kucharski, Sebastian Funk, and Rosalind M. Eggo, "The COVID-19 Response Illustrates That Traditional Academic Reward Structures and Metrics Do Not Reflect Crucial Contributions to Modern Science," *PLoS Biology* 18, no. 10 (October 16, 2020): e3000913, https://doi.org/10.1371/journal.pbio.3000913.

Early-career scholars are acutely aware of the pressures and incentives acting on professional survival, at a career stage that is frequently referred to as a bottleneck with gendered consequences.[4] Those working on COVID-19 have had to balance the urgency of the context and public good, on the one hand, with an apparently increasingly competitive job market, constant warnings about the finances of higher education institutions, and hiring freezes in UK universities, on the other. At moments, especially early on, well-meaning colleagues have cautioned that our decisions may not be strategic for career progression, and we had to ask ourselves hard questions about the sustainability of our contributions in the long and medium term.

Politicization and Communication of Research Findings

The political context in which we find ourselves contributing knowledge about COVID-19 policies in Brazil has been an ongoing source of anxiety. The pandemic has been politicized across the globe, affecting foreign affairs, global trade, and elections, to mention but a few areas of tension. In Brazil, the lack of a national coordination to fight COVID-19 has created disputes across different levels of government, and such disputes have generated contradictory and incomplete information for society. Two health ministers were sacked in the first six months of the pandemic, demonstrating that divergences regarding the seriousness and the appropriate courses of action were highly charged in Brazil's corridors of power. Because of the contentiousness of these topics at the national level, local authorities gained greater autonomy to decide how to manage the pandemic. But, in a year when local elections are held, mayors had to balance strict policies to control local case numbers and investment in health resources, with more proven means and investments to boost their popularity. In such a context, data availability, description, and analysis have been politically contentious.

Our days have been peppered with media queries. In a politically fraught environment, we have found that our data and findings have sometimes been presented in the press in a confrontational manner, in opposition to politicians' statements. For example, Andreza gave interviews to newspaper and television sources in Brazil, to which Brazil's Ministry of Health was later offered a right to respond. The ethical dilemma in such cases is that research groups, especially those publicly funded, fear budget cuts. In surveys that directly interview mayors, future contribution may be harmed if politicians feel the research group is biased.

Moreover, occasionally politicians and political parties have used our data out of context for their own purposes, placing us in the difficult position of whether or not to respond to inflammatory exchanges on social media. At one point early on, for example, the BJP party in India drew on that country's maximum policy stringency score in the international OxCGRT dataset to imply a perfect response. In anticipation of press coverage of panel survey data for Brazil, Anna sought advice from senior colleagues about whether to hold back politically sensitive information relating to citizens' opinions of Bolsonaro's handling of the crisis, in an effort to encourage a media focus on public health findings. There are no easy answers to these questions. We have been learning on the job.

Finally, we have both struggled with the fact that press coverage has tended to emphasize the Oxford brand, underrepresenting the contribution of our collaborators elsewhere. At worst, collaborators can get left out of coverage entirely, leaving our partners rightly feeling disheartened. This was part of the reason why Anna decided to create videos to accompany the OxCGRT second Brazil working paper. Media coverage also left out Andreza's partner institutions because, on television, each second matters and, unable to mention all names, priority was given to Oxford. This does not make collaborations easy in the long run.

---

[4] Annalisa Murgia and Barbara Poggio, eds., *Gender and Precarious Research Careers: A Comparative Analysis* (London: Routledge, 2019).

Conclusion

Doing research in times of COVID-19 has been a phenomenal learning experience in different ways and overall has been positive for us both. Despite worries and bumps along the way, we feel privileged that our work has been valued by the community of scientists, journalists, and the Brazilian community at large. Yet, in a COVID and post-COVID job market, we have not completely placated our own fears, and those for scholars with whom we have collaborated. Current appreciation, years hence, may not be sufficiently sticky to translate into recognition that is valued on the academic job market.

In that sense, we hope that COVID-19 catalyzes the community of scholars to broadly reevaluate the contribution of timely and accessible information sharing, which includes not only datasets but also highly informative blogs, Twitter accounts, and web pages. The current heavy emphasis on journal articles, which require editorial approval, two or more reviews, rewriting and resubmission—all processes slowing down during the pandemic—is perhaps excessive and insufficiently captures the diverse ways in which early-career scholars have provided reliable and necessary knowledge to the research community in 2020. We acknowledge our privilege in working in a highly research-focused and well-known university, which has without doubt smoothed for us some of the complex and difficult incentives that we've encountered. This acknowledgement makes us additionally concerned for talented scholars in other institutions who have invested long hours in projects.

This article has argued that scholarly activities that meet urgent policy needs do not necessarily count on the conventional academic scoreboard, which may undermine their production in the long run. Now, a full year after a new coronavirus was reported, the value of information sharing has gained attention in academia, be that via twitter accounts or blogs, and the number of academics engaging in outreach channels throughout the pandemic has been considerable. Those channels had mainly been used, pre-pandemic, to share books and articles results. We encourage readers to engage in discussions of new ways of track and value this information, and to consider strategies to support researchers conducting high-impact research and data generation in highly politicized and fast-changing fields. //