DATA&
SOCIETY

An anthology edited by
Jenna Burrell, Ranjit Singh,
and Patrick Davison

KEYWORDS OF THE DATAFIED STATE

CONTENTS

CONTRIBUTORS

CONTRIBUTORS

CONTRIBUTORS

1. Bureaucracy

Jennifer Raso is an Assistant Professor at McGill University's Faculty of Law, where she teaches Administrative Process; Science Technology and Law; and Law and Poverty. Her research investigates the relationship between discretion, data-driven technologies, and administrative law. She is particularly intrigued by how humans/non-humans collaborate and diverge as they produce institutional decisions, and the consequences for procedural fairness and substantive justice.

Victoria Adelmant is the Director of the Digital Welfare State and Human Rights Project at the Center for Human Rights & Global Justice and an Adjunct Professor at NYU School of Law. Her research and practice focuses on how the digital transformation of the state impacts human rights, particularly examining sectors that affect marginalized groups — such as welfare and immigration.

2. Capture Corporations

Burcu Baykurt is an Assistant Professor of Media Studies at the University of Massachusetts Amherst. Her research examines how digital infrastructures reshape and perpetuate durable inequalities. She is currently working on a book, *Smart as a City*, based on her fieldwork in Kansas City on experiments with implementing a gigabit internet service and a smart city program.

3. Bureaucracy

Ludmila Costhek Abílio is a sociologist who researches informal, precarious, and socially invisible work. She is currently a researcher at the Institute for Advanced Studies at the University of São Paulo and a visiting professor in the sociology graduate program at the University of Campinas. She has been closely following transformations in the work of motorcycle delivery drivers for more than ten years in the city of São Paulo and more recently, in England.

Carolina Cruz is a militant of the Landless Workers' Movement, acting in the technological studies area.

4. Data Publics

Youngrim Kim is an Assistant Professor of Media Studies at Rutgers University School of Communication and Information. She examines how state institutions build and utilize digital technologies to manage complex governance challenges, specifically in the context of public health or environmental crises. Working across media studies, critical data studies, science and technology studies (STS), she investigates how crisis response technologies reconfigure state-society relations in moments of risk and uncertainty.

5. Public Administration

Amina Abdu is a PhD candidate at the University of Michigan School of Information. Located at the intersection of values, power, institutions, and law, her research explores state data collection practices, government procurement of algorithmic technologies, and the rise of "evidence-based policy" in policymaking practices. She is interested in how various groups use data-driven technologies to gain legitimacy in policy spaces and how this in turn reshapes the nature of knowledge, expertise, and values that are seen as a legitimate basis for policy interventions.

Abigail Jacobs is an Assistant Professor in the School of Information and in the Center for the Study of Complex Systems at the University of Michigan. She is interested in how the hidden assumptions in technological systems act as a type of governance; how measurement and validation can help reveal those assumptions; and what the implications are for scientific and practical implementations of those technologies (e.g., into government settings).

6. Public Interest

Anne L. Washington is a Public Interest Technologist serving as an Assistant Professor of Data Policy at the New York University Steinhardt School. Her expertise on public sector information currently addresses the emerging governance needs of data science. As a computer scientist trained in organizational ethnography, her multi-disciplinary work considers the impact of technology on society through the lens of digital record keeping.

Joanne Cheung is an artist, architect, and researcher. Her work focuses on the political economy of digital "public space" and alternative property regimes such as community land trusts, social housing, and agrarian commons. She lectures at UC Berkeley's Haas School of Business, co-organizes the New Economy Workshop, and is affiliated with the Harvard Berkman Klein Center for Internet & Society.

7. Automation

Georgia van Toorn is an Associate Investigator at the University of New South Wales (UNSW) node of the ARC Centre of Excellence for Automated Decision-Making & Society (ADM+S). Specializing in international social policy, politics, disablement, and social justice, she explores global transformations in welfare governance, with a particular focus on processes of marketisation, the commodification of social care, and the growing impact of data analytics and algorithmic decision-making in the public sector.

Christopher O'Neill is a Research Fellow at the Monash University node of the ARC Centre of Excellence for Automated Decision-Making & Society (ADM+S). He analyzes the social and operational issues arising from the deployment of automated decision-making systems with a particular interest in the development of biometric technologies such as facial recognition cameras, and their implications for conceptions of identity and governance.

Maitreya Shah is a Fellow at the Berkman Klein Center for Internet & Society at Harvard University. As a blind lawyer and technology policy researcher, he integrates perspectives from AI ethics and critical disability studies to foreground the experiences of people with disabilities with emerging technologies. He has worked in the diverse areas of regulatory technologies, digital accessibility, public accountability, and AI governance.

Mark Andrejevic is a Chief Investigator at the Monash University node of the ARC Centre of Excellence for Automated Decision-Making & Society (ADM+S). He is Professor of Media Studies in the School of Media, Film, and Journalism at Monash University. His research covers the social, political, and cultural impact of digital media, with a focus on surveillance and popular culture.

8. Digital IDs

Mardiya Siba Yahaya is digital sociologist, researcher, and community movement builder and works as a Global Community Manager, Team CommUNITY at Article 19. Working at the intersection of digital labor, gender and cybersecurity, surveillance, counterpublics and digital identities, she investigates the implications of technology-facilitated surveillance and datafied societies on minoritized communities in the global South.

Bonnita Nyamwire is the Head of the Research Department at Pollicy Uganda, where she oversees the design of research studies, survey instruments, gathering, and analysis of data into actionable insights to influence service delivery as well as reporting on the findings. Her area of specialization is social, behavioral, and public health research.

9. Public interest technology

Maria Filippelli is a data, technology, and policy professional. As a practitioner, Maria's experience includes roles as a nonprofit data director, public interest technology census fellow, and regional planner. Her work emphasizes the need for coordination between data management, tech products, and policy to ensure responsible implementation.

10. Open data

Malavika Raghavan is a lawyer with a background in policy-focused research, studying the impacts of digitisation on low-income individuals. Her current doctoral research at the London School of Economics (LSE) examines the socio-legal consequences of digitalising welfare administration, drawing on India's experience with digital ID-linked welfare transfers. She is particularly interested in the interdependencies that develop between the technological and legal requirements of digital systems, and how they influence the exercise of administrative power within the datafied state.

11. Counterpublic

Matthew Bui is an Assistant Professor at the University of Michigan School of Information, Director of the Race and Data Justice Lab, and Co-Director of the Urban Tech Collective. His research program examines issues of race, racism, and racial justice within digital and data-driven technologies, platforms, and policy, especially at the local and community-level.

Bianca Wylie is a writer with a dual background in technology and public engagement. She is a co-founder of Digital Public, a public interest digital governance firm, where she is: a teacher, a student, an advocate for increased public power in the design and use of technology, and a democracy worker.

12. Counterdata

Seyi Olojo is a PhD student in the UC Berkeley School of Information and a visiting researcher at the Weizenbaum-Institute for Networked Society in Berlin, Germany. She is broadly interested in stratification within socio-technical systems and in investigating the role of alternative epistemological paradigms in imagining safer and more humane data practices. Her research investigates the way politicized identity categories perform within machine learning practice, most notably within commercial recommender systems and machine translation.

13. Counterdata

Vanessa Massaro is an Associate Professor and Chair of the Geography Department at Bucknell University. She is committed to innovative teaching of courses in cultural and social geography, specifically incorporating community engaged pedagogy and service learning opportunities for students. Her research draws on mixed quantitative and ethnographic methods to explore the uneven, community impacts of mass incarceration in the United States.

Darakhshan J. Mir is an Associate Professor of Computer Science and the John P. and Mary Jane Swanson Professor in Engineering and the Sciences at Bucknell University. She is also a non-residential fellow at the Center for Democracy and Technology in Washington, DC. Her research interests are focused on using computing to interrogate inequalities in our world and to help envision a new justice-oriented narrative of computing that challenges these inequities through the kinds of questions we ask and think through as scholars, community-members, and teachers. Most recently, she has been involved in a trans-disciplinary project with colleagues, students, and community partners that examines the impacts of algorithmic decision-making (and its associated ideologies baked into the carceral state) on incarcerated people in Pennsylvania.

Terrell Mosley is the Re-entry and Reintegration Coordinator at Susquehanna Valley Mediation in Selinsgrove, Pennsylvania. He works with formerly incarcerated people before, during, and after their release to constructively address conflict in their personal lives and living situations, as well as other models for restorative justice-based parole and reintegration.

Nathan C. Ryan is a Professor of Mathematics at Bucknell University and an affiliate faculty member in the Latin American Studies program. He is trained as a computational mathematician and is interested in modeling complex systems. He is involved in analyzing data from the Pennsylvania Department of Corrections related to how incarcerated people are securitized and dehumanized.

14. Missing data

Alessandra Jungs de Almeida is a PhD candidate in International Relations at the Federal University of Santa Catarina (UFSC), Brazil. She is a research affiliate at the Data + Feminism Lab and a visiting scholar (pre-doctoral fellow) at the Women's and Gender Studies Program, both at MIT. Her research connects transnational activism and feminist organizing with abortion rights, violence against women, technology, peacebuilding, and feminist security studies.

Lauren Klein is Winship Distinguished Research Professor and Associate Professor in the departments of Quantitative Theory & Methods and English at Emory University. Her research brings together computational and critical methods in order to explore questions of gender, race, and justice. Klein serves as PI of the Atlanta Interdisciplinary Network and directs the Emory Digital Humanities Lab.

Catherine D'Ignazio is an Associate Professor of Urban Science and Planning in the Department of Urban Studies and Planning at the Massachusetts Institute of Technology (MIT). She is a scholar, artist/designer and hacker mama who focuses on feminist technology, data literacy, and data activism. D'Ignazio is the director of the MIT Data + Feminism Lab which uses data and computational methods to work towards gender and racial justice, particularly as they relate to space and place.

15. Indigenous Data Sovereignty

Stephanie Russo Carroll (Ahtna-Native Village of Kluti-Kaah) is an Associate Research Professor, Udall Center for Studies in Public Policy (UC) and the Native Nations Institute (NNI) in the UC; Associate Professor in the Community, Environment and Policy Department, Mel and Enid Zuckerman College of Public Health (MEZCOPH); Associate Professor, American Indian Studies Graduate Interdisciplinary Program; and Affiliate Faculty, College of Law at the University of Arizona (UA). Her research explores the links between Indigenous governance, data, the environment, and community wellness. Her interdisciplinary research network, the Collaboratory for Indigenous Data Governance, develops research, policy, and practice innovations for Indigenous Data Sovereignty.

Marisa Elena Duarte (Pascua Yaqui Tribe/Chicanx) is an Associate Professor in Justice and Social Inquiry and the Center for Indian Education at the School of Social Transformation at Arizona State University. Her research focuses on Indigenous peoples' approaches to digital technologies, from tribally-owned Internet infrastructure to activism, Indigenous digital methods, and the politics of digitization. She applies a co-creative pedagogy in her teaching that focuses on justice theory, learning technologies in Indigenous contexts, digital activism, surveillance, and Indigenous methodologies.

Max Liboiron (Michif/Red River Métis and settler) is a Professor in Geography at Memorial University of Newfoundland and Labrador, where they direct the Civic Laboratory for Environmental Action Research (CLEAR). They have played leading roles in the establishment of the field of Discard Studies (the social study of waste and wasting), the Global Open Science Hardware Movement (GOSH), and are a figure in feminist science studies, Indigenous science and technology studies, and justice-oriented science methods.

Editors

Jenna Burrell is a Data & Society Affiliate, and previously served as the organization's Director of Research. Before joining D&S, she was a Professor at the School of Information at the School of Information, University of California, Berkeley. Burrell's research focuses on how marginalized communities adapt digital technologies to meet their needs and to pursue their goals and ideals. These days she spends a lot of time thinking about ways of protecting human control and autonomy in the wake of artificial intelligence and the possibilities for democratizing technology.

Ranjit Singh is a Senior Researcher at Data & Society focussed on qualitative methods to evaluate impact of algorithmic systems at the Algorithmic Impact Methods Lab and a research fellow with Siegel Family Endowment. With a particular focus on research equity, he also helps guide the organization's commitment to equitable research practices both internally and with its external partners. Singh's work examines the everyday experiences of people subject to data-driven practices and follows the mutual shaping of their lives and their data records.

Patrick Davison is the Associate Director of Research at Data & Society and ensures the quality of all research outputs of the organization. He is a media historian whose work examines the influence of 20th century US social science on contemporary social media.

Acknowledgments

The contributions published in this collection represent a key milestone for our 2022–2024 research strategy to work on power, control, and the datafied state. When we started this project we wanted to explore how algorithms become part of and influence state functions, how trust and doubt in public sector data infrastructures are shaped — and what the implications are for democratic practice. This exploration has involved numerous conversations at Data & Society since 2022, including small group seminars organized by Jenna Burrell, a series on methods to study the datafied state put together by Tamara K. Nopper and Jenna Burrell, and a public three-part Databite series curated by Jenna Burrell, Ranjit Singh, Tamara K. Nopper, Chaz Arnett, and Rigoberto Lara Guzmán. We would like to express our deep gratitude to everyone who participated in these conversations. Their keen interest and valuable insights have not only inspired us to continue working on this project but also significantly shaped its outcomes.

The work towards this collection began with the Keywords of the Datafied State workshop. We express our heartfelt appreciation for the time and effort that the workshop participants put into providing feedback and encouraging our contributors. Putting this workshop together would not have been possible without CJ Brody Landow's incredible event management skills. Thank you, CJ! The curatorial work on this collection involved generous feedback from and supportive conversations with Ania Calderon, Jacob Metcalf, Sareeta Amrute, Aiha Nguyen, Charley Johnson, Tamara K. Nopper, and Janet Haven. Kiara Childs provided invaluable support to the process of proofreading each contribution. Producing this collection would have been impossible without the support of the Communications

and Research Management teams — Sona Rai, Chris Redwood, Eryn Loeb, Gloria Mendoza, Alessandra Erawan, and Siera Dissmore. Special thanks to Gloria Mendoza for designing the beautiful cover and layout of this collection.

Last, but certainly not the least, we would like to thank all the contributors — Jennifer Raso, Victoria Adelmant, Burcu Baykurt, Ludmila Costhek Abílio, Carolina Cruz, Youngrim Kim, Amina Abdu, Abigail Jacobs, Anne L. Washington, Joanne Cheung, Georgia van Toorn, Christopher O'Neill, Maitreya Shah, Mark Andrejevic, Mardiya Siba Yahaya, Bonnita Nyamwire, Maria Filippelli, Malavika Raghavan, Matthew Bui, Bianca Wylie, Seyi Olojo, Vanessa Massaro, Darakhshan J. Mir, Terrell Mosley, Nathan C. Ryan, Alessandra Jungs de Almeida, Lauren Klein, Catherine D'Ignazio, Stephanie Russo Carroll, Marisa Elena Duarte, and Max Liboiron — for their generous support and engagement with the process of putting this collection together. Their commitment to a global conversation on the datafied state is exemplary and we remain grateful for their expertise and camaraderie.

All citations made in this collection are publicly accessible in a Public Zotero Library on the Datafied State. Please reach out to Ranjit Singh (ranjit@datasociety.net) if you're interested in curating and building it further.

INTRODUCTION INTRODUCTION

INTRODUCTION

REIMAGINING THE STATE IN A DATA-DRIVEN WORLD

By Jenna Burrell and Ranjit Singh

INTRODUCTION

INTRODUCTION INTRODUCTION

INTRODUCTION: REIMAGINING THE STATE IN A DATA-DRIVEN WORLD

By Jenna Burrell and Ranjit Singh

In a society of algorithms,[1] governments can become the biggest customer and the strongest guardrail for data-driven technologies. This duality in a government's relationship with technology companies has come to increasingly shape the emergent nature of the datafied state. On the one hand, the states have come to treat the tech industry as a partner. In the United States, this partnership has emerged under conditions of neoliberalism which has held bipartisan appeal for decades.[2] In the majority world,[3] this partnership has taken shape within the master narrative of modernization and progress — using computing and datafication as symbols of socioeconomic development.[4] On the other hand, the government in its role as a regulator has confronted companies that have monopolized most of the traffic on the internet. The patterns in the growth of the data economy in the last few decades shows how the attention of citizens can be commodified[5] as data and then processed to extract immense value.[6] To contend with these developments, the European Union has taken a more adversarial position toward tech monopolies and passed the most significant regulations to safeguard

1 Jenna Burrell and Marion Fourcade, "The Society of Algorithms," *Annual Review of Sociology* 47, no. 1 (July 2021): 213–37, https://doi.org/10.1146/annurev-soc-090820-020800.

2 Elizabeth Popp Berman, *Thinking like an Economist: How Efficiency Replaced Equality in U.S. Public Policy* (Princeton: Princeton University Press, 2022).

3 Shahidul Alam, "Majority World: Challenging the West's Rhetoric of Democracy," *Amerasia Journal* 34, no. 1 (January 2008): 88–98, https://doi.org/10.17953/amer.34.1.l3176027k4q614v5; Sareeta Amrute, Ranjit Singh, and Rigoberto Lara Guzmán, "A Primer on AI in/from the Majority World: An Empirical Site and a Standpoint" (New York: Data & Society Research Institute, September 14, 2022), https://datasociety.net/library/a-primer-on-ai-in-from-the-majority-world/; Ranjit Singh, Rigoberto Lara Guzmán, and Patrick Davison, eds., *Parables of AI in/from the Majority World* (New York: Data & Society Research Institute, 2022), http://dx.doi.org/10.2139/ssrn.4258527; Paola Ricaurte Quijano, "Ethics for the Majority World: AI and the Question of Violence at Scale," *Media, Culture & Society* 44, no. 4 (May 1, 2022): 726–45, https://doi.org/10.1177/01634437221099612.

4 Linnet Taylor and Dennis Broeders, "In the Name of Development: Power, Profit and the Datafication of the Global South," *Geoforum* 64 (August 2015): 229–37, https://doi.org/10.1016/j.geoforum.2015.07.002.

5 Thomas H. Davenport and John C. Beck, *The Attention Economy: Understanding the New Currency of Business* rev. ed. (Boston: Harvard Business Review Press, 2002); Morten Axel Pedersen, Kristoffer Albris, and Nick Seaver, "The Political Economy of Attention," *Annual Review of Anthropology* 50, no. 1 (2021): 309–25, https://doi.org/10.1146/annurev-anthro-101819-110356.

6 Nick Couldry and Ulises Ali Mejias, "The Decolonial Turn in Data and Technology Research: What Is at Stake and Where Is It Heading?," *Information, Communication & Society* 26, no. 4 (November 2021): 1–17, https://doi.org/10.1080/1369118X.2021.1986102.

competition in the data economy within their borders with uneven global implications.[7] A separate significant intervention has been the efforts of majority world countries[8] and indigenous states[9] in claiming sovereignty over their peoples' data, regardless of where it is held and by whom.

In writing the introductory paragraph, we made a slight change in terminology over the first two opening sentences: from government to state. This shift is crucial. State is a conceptual frame used to broadly articulate practices of governing a community of persons living on a definite territory. This element of territoriality has often implied a deep concomitant relationship between the state and the nation,[10] and that nationalism is an essential feature of identity-formation that makes up the state.[11] Of course, borders are one of the many ways of demarcating relations between people and practices of governance can also take on transnational forms,[12] such as the European Union. Yet as Begoña Artexaga succinctly articulates, "The state should … be thought of in ways that are not necessarily totally dislodged from the nation but neither attached to it."[13] Nation-state, however, is not the only analytical frame that can be used to unpack the nature of the state. The formation of any community relies on its peoples' commitment to follow its governance structures. When we become a part of a community, we also become a part of the state that is grounded in its practices of governance. A government is also made up of people who represent this community. It has a defined organization, usually codified in the form of a constitution adopted by the state. As a collective of representatives, the government is obligated to exercise the power of the state in the interest of the community that constitutes it. While the state is a shorthand to encompass a community, a government comprises those who make the rules on how to live within this community's variously constituted borders.

7 Payal Arora, "General Data Protection Regulation — A Global Standard? Privacy Futures, Digital Activism, and Surveillance Cultures in the Global South," *Surveillance & Society* 17, no. 5 (December 2019): 717–25, https://doi.org/10.24908/ss.v17i5.13307.

8 Begoña Aretxaga, "Maddening States," *Annual Review of Anthropology* 32, no. 1 (2003): 393–410, https://doi.org/10.1146/annurev.anthro.32.061002.093341.

9 Tahu Kukutai and John Taylor, *Indigenous Data Sovereignty: Toward an Agenda* (Canberra: Australian National University Press, 2016). See: https://docs.google.com/document/d/1cdj2rbU0XV14E2PJWwpxt-kvVeaxXEa-P8JBX0wBjI0U/edit?usp=sharing for clarity.

10 Begoña Aretxaga, "Maddening States," *Annual Review of Anthropology* 32, no. 1 (2003): 393–410, https://doi.org/10.1146/annurev.anthro.32.061002.093341.

11 Benedict Anderson, Imagined Communities: *Reflections on the Origin and Spread of Nationalism*, rev. ed. (New York: Verso, 2016).

12 Saskia Sassen, "Bordering Capabilities Versus Borders: Implications for National Borders," *Michigan Journal of International Law* 30, no. 3 (2009): 567–97, https://repository.law.umich.edu/mjil/vol30/iss3/2/.

13 Aretxaga, "Maddening States," 398.

This short detour into exploring the state and government conceptually offers a preview into the overall focus of this collection. Words matter. They matter because they often become a key to understanding practices. Keywords for any practice are words that may have broad or generic meanings but take on a certain specificity within the context of that practice. The practices of datafication in organizing the state are no different. By conducting interviews with government officials, being embedded in a government agency, or scrutinizing government documents or datasets, we come to better understand keywords that underlie the practices of infrastructuring the datafied state.

Keywords often cross over from institutional practices and make their way into the discourse of scholarship. Raymond Williams's classic book *Keywords: A Vocabulary of Culture and Society* documented the etymological evolution of words as markers of key shifts in culture and society.[14] Such terms capture our attention via their ambiguity, polysemy, or new frequency of use. Williams's work has spawned many other keywords collections.[15] Likewise, institutions and the public adopt language from scholarship. For example, the word "algorithm" has in the past decade gone from an arcane technical term taught and used by computer scientists to one used in mainstream media and pop culture — and invoked regularly within government as well.[16]

In our efforts to showcase how interpretive flexibility[17] manifests in keywords, this collection differs from traditional collections in significant ways.

First, keyword collections often present a single definition for each keyword with the author drawing from a vast array of scholarship to illustrate diversity in its meaning. However, the end result is a definition meant to be, more or less, comprehensive. In this collection, for several keywords we

14 Raymond Williams, *Keywords: A Vocabulary of Culture and Society* rev. ed. (New York: Oxford University Press, 1983).

15 Andrea Cornwall, "Buzzwords and Fuzzwords: Deconstructing Development Discourse," *Development in Practice* 17, no. 4/5 (August 2007): 471–84, https://www.jstor.org/stable/25548244; Craig Jeffrey and John Harriss, *Keywords for Modern India* (Oxford: Oxford University Press, 2014); Stephanie Nohelani Teves, Andrea Smith, and Michelle Raheja, eds., *Native Studies Keywords* (Tucson: University of Arizona Press, 2015); Benjamin Peters, ed., *Digital Keywords: A Vocabulary of Information Society and Culture* (Princeton: Princeton University Press, 2016); *The Keywords Feminist Editorial Collective, ed., Keywords for Gender and Sexuality Studies* (New York: NYU Press, 2021); AI Now Institute, "A New AI Lexicon: Responses and Challenges to the Critical AI Discourse," 2021, https://ainowinstitute.org/series/new-ai-lexicon; Nanna Bonde Thylstrup et al., eds., *Uncertain Archives: Critical Keywords for Big Data* (Cambridge: The MIT Press, 2021).

16 Tarleton Gillespie, "The Relevance of Algorithms," in *Media Technologies: Essays on Communication, Materiality, and Society*, eds. Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (Cambridge: MIT Press, 2013), 167–93, http://ieeexplore.ieee.org/document/6733906.

17 Trevor J. Pinch and Wiebe E. Bijker, "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other," *Social Studies of Science* 14, no. 3 (August 1984): 399–441, https://doi.org/10.1177/030631284014003004.

invited multiple teams of authors to define the same word, including public interest, bureaucracy, and counterdata. We did this to showcase how authors foreground different themes and examples to define the same keyword. We do this to disrupt the assumption that these words could or should have one single, conclusive definition.

Second, in the same spirit, we invited contributors to indicate which of the three viewpoints they planned to take in defining their keyword.

- *High-level viewpoints:* for conceptual clarification and etymological histories. This can include terms with ambiguous, multiple, or shifting definitions that are important for understanding the datafied state.
- *Viewpoints from within the datafied state:* for terms with currency and value within government.
- *Viewpoints from outside of the datafied state:* for terms that represent a critique of government generally, the datafication of government specifically, or that are used to argue for alternatives.

The first viewpoint is the more traditional one, representing the way Williams's *Keywords* collection[18] defined the genre. By explicitly including the second and third viewpoints, however, we hoped to open a door for those who find literature reviews to be a strange or inaccessible idea. Those who know what they know from being on the ground: working within government or from an outside standpoint resisting it.

Third, we encouraged academic contributors to collaborate with a first-time or non-traditional coauthor[19] who brings a distinct viewpoint, lived experience, or deeper grounding in the keyword and to think mindfully about

18 Williams, *Keywords*.

19 Mariolga Reyes Cruz, "What If I Just Cite Graciela? Working Toward Decolonizing Knowledge Through a Critical Ethnography," *Qualitative Inquiry* 14, no. 4 (June 2008): 651–58, https://doi.org/10.1177/1077800408314346.

citational justice[20] in their writing. We believed that doing so would help uplift more voices in the space of scholarship around the datafied state and expand our shared community.

Finally, this collection represents views that span the globe to highlight that the datafied state does not have a singular form. While data-driven systems as a distinct form of authority, discourse, and action have the capacity to shape the political culture of a nation-state, the state often has its own repertoire of norms, institutions, and traditions that push back.[21] The interplay between the two manifests in different meanings of a keyword in different geographies. Attending to this difference is crucial for the global project of mapping the ongoing datafication of the state.

In the following sections, we dive deeper, engaging in some definitional work to situate our readers and synthesizing the contributions in this collection to guide readers. We expand on our definition of the state by asking, "What does it mean for the state to be datafied?" We point to recurring terms like "data" and "public" that appear in multiple contributions and terms that surreptitiously found their way into several entries — "surveillance" being the most notable. We conclude with reflections on who this collection is for, what life we hope for it to have as we release it into the world, and the possible trajectory of future efforts.

What Does It Mean for a State to Become Datafied?

A state doesn't exist without the community of people it circumscribes, hence counting people has always been a constitutive element of making up

20 The Citational Justice Collective et al., "Citational Justice and the Politics of Knowledge Production," *Interactions* 29, no. 5 (August 2022): 78–82, https://doi.org/10.1145/3556549; Angela Okune, "Self-Review of Citational Practice" (Zenodo, May 21, 2019), https://doi.org/10.5281/zenodo.3066861.

21 For examples of efforts to map socio-technical change through the mutually constitutive relationship between the social shaping of technology and the technical building of nation-states, see Ranjit Singh, "Give Me a Database and I Will Raise the Nation-State," *South Asia: Journal of South Asian Studies* 42, no. 3 (May 2019): 501–18, https://doi.org/10.1080/00856401.2019.1602810; Wiebe E. Bijker, "Dikes and Dams, Thick with Politics," *Isis* 98, no. 1 (2007): 109–23, https://www.journals.uchicago.edu/doi/abs/10.1086/512835?journalCode=isis; Sheila Jasanoff and Sang-Hyun Kim, *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (Chicago: University of Chicago Press, 2015); Margaret Jack and Seyram Avle, "A Feminist Geopolitics of Technology," *Global Perspectives* 2, no. 1 (June 2021): 24398, https://doi.org/10.1525/gp.2021.24398.

the state. Building on these practices of counting, the development of the field of statistics by the beginning of the 19th century transformed the conception of the nation-state. German thinkers and statesmen of the time "brought to full consciousness the idea that the nation-state is essentially characterized by its statistics."[22] In fact, statistics as a keyword was initially employed to describe a "science dealing with the facts of a state."[23] Its importance was framed by articulating its relationship with history — "History is ongoing statistics, statistics is stationary history."[24]

The modern state has always been datafied; it is constructed through numbers and data.

Datafication of the state tends to take on an ominous form when thinking through sociological definitions of the state. Max Weber's definition is a case in point. In "Politics as a Vocation," he defined the state as, "a human community that (successfully) claims the *monopoly of the legitimate use of physical force* within a given territory."[25] The state "is considered the sole source of the 'right' to use violence."[26] Yet Weber also obliquely mentions that force is not the only means of the state, simply that it is its defining feature. Another of Weber's major contributions to our understanding of the state is his definition of bureaucracy as an impersonal system of rules that exists independently of any particular government worker who may be tasked with enacting it. This impersonality is a move toward fairer treatment of citizens,[27] although rule-following is rarely straightforward and often involves arbitrary forms of judgment.[28] While Jennifer Raso and Victoria Adelmant's contribution on bureaucracy notes the derogatory sense underlying the term's use in everyday speech today, Weber's work lacks such connotations. On the contrary, he articulated this impersonality and the calculability of

22 Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990), 18.

23 Stuart Woolf, "Statistics and the Modern State," *Comparative Studies in Society and History* 31, no. 3 (July 1989): 590, https://doi.org/10.1017/S0010417500016054.

24 Hacking, *The Taming of Chance*, 24.

25 Max Weber, *From Max Weber: Essays in Sociology*, ed. Hans Heinrich Gerth and C. Wright Mills (New York: Oxford University Press, 1946), 78, emphasis in original.

26 From Max Weber, 78.

27 "When fully developed, bureaucracy also stands, in a specific sense, under the principle of sine ira ac studio. Its specific nature, which is welcomed by capitalism, develops the more perfectly the more the bureaucracy is 'dehumanized,' the more completely it succeeds in eliminating from official business love, hatred, and all purely personal, irrational, and emotional elements which escape calculation. This is the specific nature of bureaucracy and it is appraised as its special virtue." Weber, 215–16, emphasis in original.

28 Akhil Gupta, *Red Tape: Bureaucracy, Structural Violence, and Poverty in India* (Durham: Duke University Press, 2012); Matthew S. Hull, *Government of Paper: The Materiality of Bureaucracy in Urban Pakistan* (Berkeley: University of California Press, 2012).

rules that underlie a bureaucracy's operation as its "purely technical superiority over any other form of organization. The fully developed bureaucratic mechanism compares with other organizations exactly as does the machine with the non-mechanical modes of production."[29] We can take either of Weber's definitions and consider how datafication redirects or otherwise alters the state.

Datafication can be interpreted simultaneously as (1) an amplification[30] of the state as a force and (2) as an investment in the technical superiority and impersonality of bureaucracy in organizing the state.

Approaching datafication as amplifying the state as a force raises a critical question, who is this force directed toward and against? Perhaps force by the state is desirable if it is directed at some unelected power, like the tech industry, and is wielded on behalf of a public or a marginalized group who lacks power. Yet, it is often the case that a datafied state is one in which the state and private firms link up together in deeper alignment. Neoliberal policies have often enacted this sort of model, particularly in the United States. Firms today provide services to the state that help to expand its reach and ability to oversee all people within its territorial boundaries, at its borders, and even those beyond its borders.[31] This is an alignment in which tech firms function as capture corporations (as Burcu Baykurt argues in her contribution). The force of the state is not one opposed to private tech, but interlinks with firms to exert force and control over the populace. This alignment is also a way of excising parts of the state, reducing costs and gaining efficiencies, but also making the state less able to uphold public values. Ludmila Costhek Abílio and Carolina Cruz note this in their own contribution on bureaucracy, showing how platform companies have taken on certain functions and services that traditionally belonged to the state.

29 Weber, *From Max Weber*, 214.

30 Philip E. Agre, "Real-Time Politics: The Internet and the Political Process," *The Information Society* 18, no. 5 (October 2022): 311–31, https://doi.org/10.1080/01972240290075174; Kentaro Toyama, *Geek Heresy: Rescuing Social Change from the Cult of Technology* (New York: PublicAffairs, 2015), 17–37.

31 John Cheney-Lippold, "Jus Algoritmi: How the National Security Agency Remade Citizenship," *International Journal of Communication* 10 (2016): 1721–42, https://ijoc.org/index.php/ijoc/article/view/4480.

In some cases, the populace may actively join this alignment between the state and private sector tech firms as well. Youngrim Kim, writing on data publics, argues that notions of patriotism and duty in Korea led citizens to join together with government agencies and tech companies to realize a national COVID-19 surveillance infrastructure. She argues it is a western conceit to assume citizens, through their membership in civil society groups, are primarily an opposing force that seeks to resist or reform the state.

If, alternatively, we understand datafication as a way of investing in the more logical, impersonal enactment of rules within bureaucracies, then the replacement of a bureaucrat with a computer certainly gives the appearance of greater impersonality, though it also serves to conceal the human traces and judgment calls underlying an automated façade. It also risks enacting rules in ways that are inflexible to the point of being nonsensical or even cruel. Amina Abdu and Abigail Jacobs's contribution on public administration makes this point that datafication became a way that public agencies (composed of unelected civil servants) sought to solidify their legitimacy;[32] however, new criticism of tech is calling this legitimacy into question. Considering both these possibilities together, it is not clear whether the datafied state always acts in the public interest. Anne L. Washington and Joanne Cheung argue that the public interest must be grounded in engaging with edge cases and those in the margins.

While the datafied state may rhetorically aspire to uphold the public interest, whether it does so in practice is an empirical question.

Opening up the possibilities of empirical investigation draws our attention to the everyday lived experience of the datafied state. In this respect, anthropological approaches to defining the state trace a different genealogy of the field. Anthropology has offered alternatives to the state (as a western

32 See also, Theodore Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1995).

phenomenon), including social structures that operate at a smaller scale to govern populations through the normative and institutional forms that define culture, such as religion, family, civil society, and the economy. In anthropology, this notion of the state as a western phenomenon dissolved with the firmer integration of political economy into the study of culture.[33] The state was traditionally approached in the field "as a given — a distinct, fixed and unitary entity that defines the terrain in which other institutions function."[34] However, more recently, attempts have been made to "bring together the ideological and material aspects of state construction, and understand how 'the state' comes into being, how 'it' is differentiated from other institutional forms, and what effects this construction has on the operation and diffusion of power throughout society."[35] Attending to the cultural constitution of the datafied state involves following ongoing and emergent cultural struggles that are waged in two interrelated aspects of state-formation: first, in the sphere of representation, and second, in the domain of everyday practices of bureaucracies. Writing on automation, Georgia van Toorn, Chris O'Neill, Maitreya Shah, and Mark Andrejevic illustrate both these aspects in the ongoing investments in automation by governments across the world. They showcase how automation represents speed, efficiency, and precision, and enacts a cascading logic[36] that fundamentally reorganizes everyday bureaucratic work.

Exploring these ongoing forms of reorganization further, bureaucratic procedures routinely rely on tools to manage state-citizen relations. Citizen data is one such tool. Every tool has its affordances and limits; they offer a perspective embedded in their very construction. Mardiya Siba Yahaya and Bonnita Nyamwire show in their contribution how citizen data collected during the process of issuing biometrics-based digital IDs not only represents bodies of citizens made available for scrutiny at a distance, but

33 George E. Marcus and Michael M. J. Fischer, *Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences*, 2nd ed. (Chicago: University of Chicago Press, 1999).

34 Aradhana Sharma and Akhil Gupta, eds., *The Anthropology of the State: A Reader* (Malden, MA: Blackwell Publishing, 2006), 8.

35 Sharma and Gupta, 8.

36 Mark Andrejevic, *Automated Media* (New York: Routledge, 2019).

also the state itself as a collective moving away from corruption and toward progress and development. The everyday experiences of living with data mutually shape the meaning of the state for bureaucrats and citizens alike.[37] As a public interest technology practitioner involved in negotiation over how and when to use digital tools in delivery of government services, Maria Filippelli offers us a window into these experiences.

Clues to the material and discursive formation of the datafied state lie in mundane bureaucratic procedures that must accommodate data-driven technologies.

Questioning and Resisting the Datafied State

The word "public" (as in "public sector") often serves as a proxy for the state. In some instances, it means ownership by the state. When the datafied state is a topic of research, data-driven systems owned and operated by the government and the infrastructures they are imbricated within become suitable case studies.[38] A number of terms in this collection include this word ("counterpublics"; "public administration"; "data publics"; "public interest"). However, the word "public" takes on a multitude of meanings in these contributions, moving beyond government ownership. In some other instances, public(s) emerge as manifestations of "amorphous and unarticulated" collectives of people who organize themselves in the face of problems and/or issues that affect them to express their concerns.[39] By acting upon such problems, John Dewey argued,[40] the public manifests its capacity to hold the state accountable. Finally, the use of the word "public" is also a part of exploring the principle of openness, as in the public disclosure of data. In her contribution, Malavika Raghavan highlights how despite diverse definitions

37 Bidisha Chaudhuri, "Programmed Welfare: An Ethnographic Account of Algorithmic Practices in the Public Distribution System in India," *New Media & Society* 24, no. 4 (April 2022): 887–902, https://doi.org/10.1177/14614448221079034.

38 Lina Denick, Joanna Redden, Arne Hintz, Harry Warne, "The 'Golden View': Data-Driven Governance in the Scoring Society," *Internet Policy Review* 8, no. 2 (June 30, 2019), https://policyreview.info/articles/analysis/golden-view-data-driven-governance-scoring-society.

39 John Dewey, *The Public and Its Problems* (New York: H. Holt and Company, 1927), 131.

40 Dewey, *The Public and Its Problems*.

of "open data," certain common features hold. Open data is concerned with publicly accessible datasets and the negotiation over the formats in which they are made available.

"Public" melds the institutions of government with the people subject to that government.

The state, in theory, represents the interests of all people within its boundaries. This framing inevitably brings us to the question — how unified are these interests? What about when these interests are at odds? What differences are being glossed over when "the public" is referred to with the singular "the"? Many of the contributions in this collection pick up on this theme. Washington and Cheung's public interest and Matthew Bui and Bianca Wylie's counterpublic both show how the public is rhetorically deployed to serve the interests of those in power. Both essays engage the notion of the public with necessary skepticism, investigating cases of urban public space management to illustrate who is implicitly included and excluded by the term.

"The public" is a term that holds power by implying a consensus that often does not exist.

Facing a state aligned with some publics over others, citizens respond using a diverse set of new and old tactics. The two contributions to this collection on counterdata illustrate these tactics in distinct ways. While Seyi Olojo takes the route of a broader historical review and makes a definitional intervention, Vanessa Massaro, Darakhshan J. Mir, Terrell Mosley, and Nathan C. Ryan re-examine how recidivism is measured in the context of the US criminal legal system. Practices of using data to counter policies and practices of the state have a long history. Olojo points to the work of

Ida B. Wells who collected statistics on lynching in the late 19th century to show the unjust and pervasive targeting of Black men by this form of extra-judicial violence. Similarly, sociologist W. E. B. DuBois created data visualizations to represent Black Life and to challenge monolithic and racist representations of Blackness.[41]

If data can be used to construct the state, it can also be used to deconstruct it.[42]

It should neither be surprising nor revelatory that many civil society groups and social movements have made acquiring, analyzing, and presenting data a part of their practices of seeking justice, policy change, or simply greater visibility.[43] In part, they seek the legitimacy that quantification and data have achieved in the modern state. They often leverage, as Raghavan also shows, the bounty that recent open data efforts have offered in accessing government data for the sake of transparency. However, gaps in data collection are pervasive. Alessandra Jungs de Almeida, Lauren Klein, and Catherine D'Ignazio, in their contribution on missing data, move beyond the reinterpretation of official data collected by the state to call attention to under-resourced efforts to painstakingly collect data to fill gaps created by state neglect or intentional silence on critical social problems. Data is neither raw[44] nor always available; it must be produced to become a resource for building as well as resisting the state. Data has politics[45] that are at play in the infrastructural processes of data collection, circulation, curation, and interpretation.[46]

It is against the backdrop of these processes of managing citizen data that contests are fought over whose interests the state aligns with and who gets left behind.

Finally, Stephanie Russo Carroll, Marisa Duarte, and Max Liboiron take these contests over data as a point of departure in their contribution on

41 Whitney Battle-Baptiste and Britt Rusert, eds. *W. E. B. Du Bois's Data Portraits: Visualizing Black America* (New York: Princeton Architectural Press, 2018).

42 For a more detailed account of how 'deconstruction' is used in STS studies of how technoscience is enacted in legal contexts and court proceedings through cross-examination of evidence, refer to the work of Sheila Jasanoff, *Science at the Bar: Law, Science, and Technology in America* (Cambridge: Harvard University Press, 1995), 211–15.

43 Stefan Baack, "Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism," *Big Data & Society* 2, no. 2 (December 2015): 1–11, https://doi.org/10.1177/2053951715594634; Stefania Milan and Lonneke van der Velden, "The Alternative Epistemologies of Data Activism," *Digital Culture & Society 2*, no. 2 (December 2016): 57–74, https://doi.org/10.14361/dcs-2016-0205; Emily Edwards et al., "Shaheen Bagh: Making Sense of (Re) Emerging 'Subaltern' Feminist Political Subjectivities in Hashtag Publics through Critical, Feminist Interventions," *New Media & Society*, December 7, 2021, 1–22, https://doi.org/10.1177/14614448211059121; Lauren Kogen, "From Statistics to Stories: Indices and Indicators as Communication Tools for Social Change," *The International Journal of Press/Politics*, (April 2022) https://doi.org/10.1177/19401612221094246.

44 Lisa Gitelman, *"Raw Data" Is an Oxymoron* (Cambridge: MIT Press, 2013).

45 Evelyn S Ruppert, Engin Isin, and Didier Bigo, "Data Politics," *Big Data & Society* 4, no. 2 (July 2017): 1–7, https://doi.org/10.1177/2053951717717749.

46 Ranjit Singh and Steven J. Jackson, "Seeing Like an Infrastructure: Low-Resolution Citizens and the Aadhaar Identification Project," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 2021): 1–26, https://doi.org/10.1145/3476056.

Indigenous data sovereignty to break free from certain core assumptions about the state. They show how assumptions around territoriality are grounded specifically in settler colonialism, a particular practice of state building by seizing land. The result is the dispossession of Indigenous people from land and lives, cultural artifacts, as well as knowledge. Data about Indigenous people, collected in the course of research or government demographics, likewise has often been misanalyzed and misinterpreted to uphold power; it is often used against the interests of Indigenous groups from whom it was collected. They observe that "open" data is a permissive framework that always benefits settlers and their systems of government over Indigenous communities. Their contribution includes guidelines to establish the terms of collaboration around data between Indigenous and non-Indigenous groups. They ask the more fundamental question about whether and how data can be collected, who holds it, who owns it, and who has access to it, arguing that Indigenous groups globally are not another minoritized constituency who use data to make appeals to an overarching state.

Indigenous states exist with their own sovereign systems of governing relations informed by specific intellectual, ceremonial, and ancestral traditions — including relations embedded in and through data; a sovereignty that they must continually reassert and defend against duplicitous, treaty-breaking settler governments.

Conclusion: The Search for Keywords of the Datafied State

In writing up the previous two sections on defining the datafied state, we have covered all contributions to this collection. Yet, as must be obvious to

our readers, there are many keywords that have been left behind — abolition, surveillance, procurement, border control, to name a few. We have had contributors whose lives interrupted the completion of their contributions; busy bureaucrats, public servants, and experts with deep experience in public administration couldn't be brought into the collection as successfully as we had hoped. However, we see these setbacks as normal natural challenges of taking on the task of putting together a collection of keywords for the datafied state. Our intention in working toward building this community was never to be comprehensive, rather it was to invite a broader conversation on the shifting nature of the state as it appropriates ever more complex data-driven systems. Furthermore, readers will find that many of the contributions in the collection can be read from the lens of a different keyword. Many of the contributions, for example, name surveillance[47] as a foundational concern in the transformation of state-citizen relations through data. Similarly, it is hard to separate conversations on public interest and public private partnerships from discussions on procurement.[48] Yet, our effort is partial; this collection is a product of the community that we could gather around our shared research interest, while being physically located within the United States. We hope that our readers see it as a resource for gathering their own communities to engage with the ongoing emergent challenges of contending with the datafied state and as an invitation to explore which keywords matter most to them.

47 Torin Monahan and David Murakami Wood, eds., *Surveillance Studies: A Reader* (New York: Oxford University Press, 2018); Virginia Eubanks, *Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2017); Michele Gilman and Rebecca Green, "The Surveillance Gap: The Harms of Extreme Privacy and Data Marginalization," *NYU Review of Law and Social Change* 42, no. 2 (2018): 55, https://socialchangenyu.com/review/the-surveillance-gap-the-harms-of-extreme-privacy-and-data-marginalization/; Chaz Arnett, "Race, Surveillance, Resistance," *Ohio State Law Journal* 81, no. 6 (2020): 1103–42, https://kb.osu.edu/items/70ad60c0-d30e-4e7b-b740-a7d03c0095a9.

48 Mona Sloane, Rumman Chowhury, John C. Havens, Tomo Lazovich, and Luis Rincon Alba, "Procurement as Policy: Administrative Process for Machine Learning," *Berkeley Technology Law Journal* 34, no. 3 (2019): 773–852, https://doi.org/10.15779/Z38RN30793; Mona Sloane et al., "AI and Procurement: A Primer," PDF (New York: New York University, 2021), https://doi.org/10.17609/BXZF-DF18.

BUREAUCRACY
BUREAUCRACY
BUREAUCRACY

BY

JENNIFER RASO and VICTORIA ADELMANT

BUREAUCRACY
BUREAUCRACY
BUREAUCRACY

BUREAUCRACY

By Jennifer Raso and Victoria Adelmant

Bureaucracy is the original machinery of datafication. It is an organizational form made up of people, information, rules, and technologies. Bureaucracies are designed to gather, control, curate, and rely upon information.[1] But they are ineffective without people. As an organizational form, then, bureaucracy arranges authority among the people who work within it, distributing and delegating decision-making power to different tiers of civil servants and others. Bureaucracy thereby fulfills a critical legal function as it organizes and allocates state decision-making authority. Bureaucracy also structures legal relations between the state and the public. The actors, techniques, and systems comprising bureaucracy apply legal rules to real-life situations where most people experience government. Bureaucratic actors (from public officials to decision-making software) thus profoundly impact people's lives and create, reduce, or amplify structural inequalities.[2]

The term "bureaucracy" has also long been derogatory shorthand for inefficient, impenetrable government. Specific bureaucracies, and bureaucracy more broadly, are regular targets for transformation projects that aim for a government ruled by common sense rather than tied up in red tape.[3] For decades, state officials have eagerly adopted new technologies to change how their bureaucracies function.[4] By the early 1990s, digitalization was even proclaimed a means of "ending bureaucracy."[5] Today, new data-driven tools and methods continue to be deployed as an antidote to inefficient processes.[6] State datafication thus features governments adopting ever more advanced computational tools, techniques, and systems and automating many components of decision-making processes across bureaucracies.

1 Max Weber, "Bureaucracy," in *Economy and Society* Vol. 2, eds. Guenther Roth and Claus Wittich (Berkeley: University of California Press, 1978).

2 Ruha Benjamin, *Race After Technology* (Cambridge: Polity Press, 2019); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018); Rashida Richardson, "Racial Segregation and The Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities," *Berkeley Technology Law Journal* 36, no. 3 (October 2022): 1052–1090, https://doi.org/10.15779/Z38PN8XG3v.

3 On Ontario, Canada's "Common Sense Revolution," Ian Morrison, "Ontario Works: A Preliminary Assessment,"' *Journal of Law and Social Policy* 13, (January 1998): 1–46. Alberta, Canada, which is dominated by Conservative politicians, even has a Ministry of Red Tape Reduction (which, ironically, enlarges the Province's bureaucracy).

4 Peter Crooks, "Bureaucracy," in *Information: A Historical Companion*, eds. Ann Blair, Paul Duguid, Anja-Silva Goeing, and Anthony Grafton, (New Jersey: Princeton University Press, 2021), 343–348.

5 Gifford Pinchot and Elizabeth Pinchot, *The End of Bureaucracy and the Rise of the Intelligent Organization* (Oakland: Berrett-Koehler, 1994).

6 The OECD's Digital Government Policy Framework argues that "digital government" should "overcome bureaucratic legacies," provide services that are "less bureaucratic," and should aim at transformation and redesign as opposed to the mere digitization of existing Weberian bureaucracy. OECD, *The OECD Digital Government Policy Framework: Six dimensions of a Digital Government*, OECD Public Governance Policy Papers no. 02 (2020):7–8, 29.

But *do datafied state initiatives end or extend bureaucracy?* This keyword entry explores this question in two parts. First, it reflects on who and what constitutes bureaucracy as the state is datafied, and how datafied state initiatives displace and disperse, rather than replace, the people and systems that make up bureaucracy. Second, it examines how ongoing datafication initiatives affect bureaucracy's specific legal function, or how bureaucracy organizes and applies decision-making authority. In doing so, it explores how datafied state initiatives disperse decision-making and the implications for accountability mechanisms.

Reconstituting Bureaucracy in Datafied States

The use of information management technologies, and even the phenomenon of datafication, are long-standing features of bureaucratic operations. As an organizational form, bureaucracy collects, centralizes, systematizes, and processes information. For the administrative branch of government to organize itself and exert control, its agencies must simplify social realities to more easily govern them.[7] Datafication, or abstracting the natural and social world into information forms that enable state agencies to analyze and act on that information, has therefore been performed by bureaucracies for centuries. These processes were central to empire building and governance, with the systematization of information crucial to imperial powers' control from afar.[8] These have also been the central means through which state agencies come to know and manage populations and individuals, from census tracts to passport documents.[9] Neither the datafication of the state nor the centrality and necessity of bureaucracy as an information management system are themselves new.

7 James Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998).

8 Katarzyna Cieslik and Dániel Margócsy, "Datafication, Power and Control in Development: A Historical Perspective on the Perils and Longevity of Data," *Progress in Development Studies* 22, no. 4 (October 2022): 352–373, https://doi.org/10.1177/14649934221076580.

9 Sara Dehm, "Passport" in *International Law's Objects*, ed. Jessie Hohmann and Daniel Joyce (Oxford: Oxford University Press, 2019), 342–356.

Because information collection and systematization are central to bureaucracies' work, bureaucratic actors have long adopted new methods and technologies.[10] Some of the earliest computers were themselves developed to facilitate bureaucratic information management. For instance, a device for performing statistical calculations was designed to speed up the US Census Bureau's process for tabulating statistics from the 1890 census.[11] The introduction of information and communications technology across governments starting in the 1970s thus extended this longer historical practice.

The Driving Role of Critique in Datafied State Projects

New technologies have been eagerly introduced by public administrators not only to facilitate information management but also to respond to bureaucracy's perceived shortcomings. Since the term was coined, "bureaucracy" has carried overwhelmingly negative connotations.[12] New technologies have long been promised to cure administrative inefficiencies. Indeed, the rise of computerization in the 1980s led to claims that technologies could "end bureaucracy."[13] Today, international organizations, ministerial offices, and consultancy firms still propose that governments overcome bureaucracy by further datafying and digitalizing their operations.[14] These approaches generally reify bureaucracy as a set of unnecessary processes within a bloated administration, rather than understanding it as a complex assemblage of systems, rules, and people. These accounts also conveniently overlook the fact that technologies have always been integral to bureaucracy and they may contribute to (rather than solve) the problem of inefficient processes.

10 Michael Adler and Paul Henman, "E-Justice: A Comparative Study of Computerization and Procedural Justice in Social Security," *International Review of Law, Computers & Technology* 15, no. 2 (July 2001): 195–212, https://doi.org/10.1080/13600860120070510.

11 Stan Augarten, *Bit by Bit: An Illustrated History of Computers* (Boston: Houghton Mifflin, 1984).

12 Anthony Grafton, Anja-Silvia Goeing, Paul Duguid, and Ann Blair, *Information: A Historical Companion* (New Jersey; Princeton University Press, 2021), 343.

13 Gifford Pinchot and Elizabeth Pinchot, *The End of Bureaucracy and the Rise of the Intelligent Organization* (San Francisco: Berrett-Koehler, 1994).

14 OECD, *Embracing Innovation in Government Global Trends* (OECD, 2017); Walt Bogdanich and Michael Forsythe, *When McKinsey Comes to Town: The Hidden Influence of the World's Most Powerful Consulting Firm* (New York: Doubleday, 2022); Mariana Mazzucato and Rosie Collington, *The Big Con: How the Consulting Industry Weakens Our Businesses, Infantilizes Our Governments, and Warps Our Economies* (New York: Penguin Press, 2023).

Attempts to overcome bureaucracy through digitalization rely heavily on critiques of the *people* animating bureaucracy. Officials are widely caricatured as stubbornly resisting change, sluggish, and capricious, particularly in frontline settings where their decisions affect high-stakes programs like welfare or immigration. Critics may also raise concerns about bias, corruption, and fraud. Such critiques often inspire the introduction of technologies to manage public officials, such as performance targets, workplace surveillance, and regular reporting to more senior officials. These tools continue pre-existing bureaucratic practices of governing officials through tools and techniques that hierarchically allocate and control decision-making power.

More recently, critiques have driven government agencies to adopt digitalized systems and data-driven tools to explicitly perform some decision-making tasks previously performed by human officials. Bureaucracy prioritizes the consistent application of rules, and data-driven algorithmic systems are perceived as better able to reliably implement decision-making criteria and deliver objective outputs, representing "the electronic equivalent of Weber's objective and impartial professional."[15]

This narrative is particularly pronounced in India, a crucial first mover in pioneering datafied initiatives across the administrative state and a key proponent of digital government on the international stage. Enthusiasts from across India's technology industry and government promote digitalization as a way to achieve Modi's vision of an India free from middlemen who obstruct public service delivery.[16] Leakage from welfare budgets and concerns about corruption among bureaucrats dominated early arguments in favor of a nationwide digital identification system.[17] The resulting digital ID system, Aadhaar, promised to eliminate middlemen by using technologies rather than local bureaucrats to deliver government services. Here and

15 Frank Bannister, "In Defence of Bureaucracy: Governance and Public Values in a Digital Age," *Beyond Bureaucracy: Towards Sustainable Governance Informatisation*, eds. Alois A. Paulin, Leonidas G. Anthopoulos, and Christopher G. Reddick, (Springer: Public Administration and Information Technology Vol. 25, 2017).

16 Bidisha Chaudhuri and Lion König, "The Aadhaar scheme: a cornerstone of a new citizenship regime in India?" *Contemporary South Asia* 26, no. 2 (2018): 127–142, https://doi.org/10.1080/09584935.2017.1369934.

17 Nandan Nilekani, *Imagining India: The Idea of a Renewed Nation* (New York: Penguin, 2009). For a critique of this narrative, see Reetika Khera, "Impact of Aadhaar in Welfare Programmes," in *Dissent on Aadhaar: Big Data Meets Big Brother*, ed. Reetika Khera (Telangana: Orient BlackSwan, 2019).

elsewhere, data-driven systems are presented as a means of replacing human administrators.

But datafication fails to eliminate bureaucratic problems and replace human decision-makers. Instead, it expands bureaucracy beyond conventional civil servants and government offices and across a wider range of sites and actors. In the process, datafication may exacerbate old bureaucratic problems as well as introduce new ones.[18]

Displacement, Not Replacement

Technologies have long been portrayed as replacing human officials. When personal computers were introduced across governments 50 years ago, officials no longer had to rely on clerks and librarians physically searching for and retrieving information. However, the task of managing and maintaining informational infrastructures did not disappear; rather, it was dispersed to other actors — from those tasked with data entry to those maintaining software.

Today, technologies may appear to complete most tasks previously performed by frontline officials. For instance, software may automatically determine applicants' eligibility for some public programs. People might apply for government services online. But even when systems offer a "digital-only" experience, datafied state initiatives *displace* rather than replace the human labor needed to keep bureaucracies functioning. Critical work is spread out among a multitude of actors, as frontline workers are joined by program users, librarians, nonprofit and charity workers, tech designers, programmers, and many others. Each of these actors performs vital data entry, system maintenance, and even decision-making tasks, and administrative burdens

18 Anumeha Yadav, "Reporting the World's Largest Biometric Project," *Lives of Data: Essays on Computational Cultures from India* (Amsterdam: Institute of Networked Cultures, 2020); Marion Fourcade and Jeffrey Gordon, "Learning Like a State: Statecraft in the Digital Age," *Journal of Law and Political Economy* 1, no.1 (2020): 78–108, https://doi.org/10.5070/LP61150258.

move around with crushing and freeing effects.[19] This dispersal requires a wider lens to examine bureaucracy in a datafied state, one that encompasses the many new entities that act as data providers, co-deciders, and system designers.

For example, the digital "self-service" model of the Universal Credit welfare program in the United Kingdom relies heavily on the labor of data providers. These actors include benefits claimants, as well as employers, software applications, and caseworkers. In the process, the administrative burdens on claimants and other data providers may increase. To file a claim, for example, an applicant must enter personal information into an online form and verify her identity online. Her digital account will then be activated with a to-do list requiring her to book a caseworker interview: only then will she meet with an official in person. The claimant will send and receive messages through an online journal as her primary communication channel. Many of the messages she receives will be automated, others will be created by staff in service centers across the country, and some will be written by her caseworker.[20] Universal Credit software will use employer-provided data held by the tax office to calculate her monthly benefit eligibility and payments. This digital avenue almost entirely replaces alternative methods of interacting with the welfare system: it is "certainly not possible to make such a claim by turning up at a [welfare] office … and handing in a paper claim form."[21] This scenario, of course, may generate exclusion and exacerbate social inequalities. While digital systems work well for some, those who are most likely to face difficulties in filing and managing claims through online systems are also more likely to need access to welfare programs.[22] Some marginalized individuals, for instance, may find online claims and communication through an online journal to be far more challenging than paper forms and in-person channels, particularly if their experiences fail to

19 Pamela Herd and Daniel Moynihan, *Administrative Burden: Policymaking by Other Means* (New York: Russell Sage, 2019); Michael Lipsky, "Bureaucratic Disentitlement in Social Welfare Programs," *Social Service Review* 58, no. 1(March 1984): 3–27, https://doi.org/10.1086/644161.

20 Richard Pope Consulting Ltd. "Universal Credit: Digital Welfare." London: Part Two Digital, April 2020. https://pt2.works/files/universal-credit.pdf.

21 *GDC v Secretary of State for Work and Pensions (UC)* [2020] UKUT 108 (AAC), 8.

22 Sophie Howes and Kelly-Marie Jones, "Computer Says 'No!' Stage Two: Challenging decisions," *Child Poverty Action Group*, (July 2019), https://cpag.org.uk/sites/default/files/files/policypost/Computer%20says%20no%21%202%20-%20for%20web.pdf.

correspond to preset options in digital forms. Administrative burdens can intensify for these individuals.

Digital systems also *disperse* administrative burdens. Not only are claimants, as data providers, tasked with completing and continually updating digital forms, but the burden also spreads outward to others who must also submit data. Employers, for example, are responsible for providing data about claimants' wages — a responsibility with serious consequences, as erroneous or incomplete data can lead to claimants' benefits being suspended.

In other settings, data-providing tasks are spread so widely that frontline bureaucrats appear to vanish entirely. In Norway, for example, child benefits are distributed automatically: rather than claimants completing a form and caseworkers receiving and approving the application, software uses information in government databases to identify recipients and disburse child benefits without any role for caseworkers (or claimants) at all.[23] Likewise, the US and Australian governments have drawn on tax and benefits data to automatically generate debt notices, many of which are erroneous.[24] Similar initiatives are being introduced by immigration and border security agencies. While these initiatives may shift burdens away from frontline officials and members of the public at the moment when a benefit is granted or a debt is created, they also defer administrative burden into the future wherever a data provider — such as an employer, landlord, doctor, etc., — generates suspect data (i.e., data that challenges one's eligibility for a particular benefit or status or suggests that a debt may be owed). In these situations, procedural justice inverts. Members of the public whose data suggests that they have received higher benefits payments than they were eligible for or are at risk of overstaying a visitor visa, for instance, must then prove the opposite, often with insufficient information about why they have been flagged as debtors or risky subjects in the first place.

23 Karl Kristian Larsson, "Digitization or Equality: When Government Automation Covers Some, but Not All Citizens," *Government Information Quarterly* 38, no. 1 (January 2021): 1–10, https://doi.org/10.1016/j.giq.2020.101547; Hendrik Scholta et al., "From One-Stop Shop to No-Stop Shop: An E-government Stage Model," *Government Information Quarterly* 36, no. 1 (January 2019): 11–26, https://doi.org/10.1016/j.giq.2018.11.010.

24 Terry Carney, "Bringing Robo-debts Before the Law: Why It's Time to Right a Legal Wrong," *Law Society NSW Journal*, (August 2019), https://lsj.com.au/articles/why-robo-debt-bringing-robo-debts-before-the-law-why-its-time-to-right-a-legal-wrong/; Doaa Abu Elyounes, "'Computer Says No!': The Impact of Automation on the Discretionary Power of Public Officers," *Vanderbilt Journal of Entertainment & Technology Law* 23, no. 3 (2021): 451–516.

The work of co-deciders is also dispersed and transformed, rather than eliminated. Digital tools may automate some of the processes for which public officials were previously responsible, assume part of their role, and change how decisions are produced. As drones, risk assessments, and biometric data collection and processing tools become integral to border administration, fingerprint and iris scans strongly influence (and even co-create) border agency decisions. Border officials may conceptualize these tools as sources of evidence, though the tools and officials together decide which fingerprints and scans are acceptable and whether they are more reliable than the statements of travelers seeking to cross the border.[25] As for welfare programs, software may calculate benefits and generate decision letters, but these outputs depend on how caseworkers characterize their data inputs about benefits applicants.[26] Bureaucratic decision-making tasks are thus more widely distributed between tools and people.

Access to a human co-decider may be an advantage in some settings and a disadvantage in others. In the Universal Credit program, reaching a human official is a sought-after privilege because of the infrastructural barriers to claimants directly connecting with officials, such as their online journal and overwhelmed call centers. In immigration, by contrast, the apparent absence of human officials may signal privilege. For example, in Canada's automated visa approval program, an algorithmic tool sorts through tourist visa applications from Chinese and Indian citizens and automatically approves applicants with "low-risk" characteristics. Only those applications the tool flags as suspect are reviewed by a human official.[27] Here, the datafied state may be frictionless for low-risk applicants but onerous for higher-risk applicants who may puzzle over why algorithmic and human co-deciders denied their visa application.

25 Petra Molnar, "Territorial and Digital Borders and Migrant Vulnerability Under a Pandemic Crisis," in *Migration and Pandemics: Spaces of Solidarity and Spaces of Exception*, ed. Anna Triandafyllidou, (Springer: IMISCOE Research Series, 2022), 45–64.

26 Jennifer Raso, "Displacement as Regulation: New Regulatory Technologies and Front-Line Decision-Making in Ontario Works," *Canadian Journal of Law & Society* 32, no.1 (June 2017): 75–95.

27 Canada, Immigration Refugees and Citizenship Canada, *Algorithmic Impact Assessment – Advanced Analytics Triage of Visitor Records Applications* (2022) https://open.canada.ca/data/en/dataset/01396e33-2c69-47e5-9381-32e717943b96.

Finally, datafied state initiatives more clearly distribute bureaucratic tasks to system designers. Just as archival tasks have shifted to the creators and maintainers of digital databases, datafied state projects rely on system designers' expertise. In 2001, Jane Fountain documented "the growth in the number of technical analyst positions required to develop, program, maintain, and service increasingly information-based federal bureaucracies" and noted these analysts' growing dominance.[28] As bureaucracies become ever more datafied, different actors and forms of knowledge — "user designers," coders, and data analysts — become more central to bureaucratic functioning.[29] As a result, funds are spent on a broader set of actors to build and operate digital government, which also (re)directs resources from government officials and program beneficiaries toward engineers, designers, and consultants.

Thus, the people and systems that constitute bureaucracy stretch far beyond conventional civil servants and government offices. Decision-making power has always been diffused by bureaucratic arrangements, but datafied state initiatives spread it out even more widely. For example, when software co-creates eligibility decisions, its designers (plus the many others identified above) become crucial bureaucratic actors. Because these designers shape how a software's digital interface operates, they influence how administrative agencies are accessed and experienced.[30] It then becomes vital for us to explore how decision-making power operates among these distinct, diffuse components of bureaucracy in the datafied state.

28 Jane Fountain, *Building the Virtual State: Information Technology and Institutional Change* (Washington D.C.: Brookings Institution Press, 2001), 62.

29 Adelmant, Victoria, and Joe Tomlinson. "Who Builds Digital Government?: Accountability in the Private Sector's 'Agile' Reconstruction of the Administrative Justice System." *Public Law*, no. 2 (April 2023): 196–206, https://doi.org/10.3316/agispt.20230721091866.

30 Adelmant and Tomlinson, "Who Builds Digital Government?"

Changes to the Legal Function(s) of Bureaucracy

As they reconfigure frontline work and spread decision-making authority, datafied state initiatives disrupt bureaucratic hierarchies and practices. These features do more than simply streamline data collection and storage; they are also crucial for conventional legal accountability mechanisms. Disrupting them creates two related challenges for bureaucracy's legal operations. First, functionally speaking, it becomes even more difficult for people within and outside of the bureaucracy to understand how decisions are produced and who (or what) is responsible. Second, these functional issues make it exceptionally tricky to ensure that decision-making institutions meet minimum accountability standards.

Practical Opacity

The architects of government digitalization initiatives often intend to disrupt bureaucratic hierarchies as they overcome bureaucracy. But because digitalization disperses decision-making authority, it creates substantial practical challenges. For instance, technical glitches often lead to widespread bureaucratic errors affecting tens of thousands of people. These people may be affected by an incorrect decision. A routine software update incompatible with Apple's operating system might spur a border crossing app to erroneously order thousands of travelers to quarantine, for example.[31] When such errors arise, officials may be infrastructurally barred from fixing the issue themselves.

Distributed decision-making power also makes it exceedingly difficult for people affected by a decision that seems blatantly wrong to know who, or what, is responsible for that decision and where to seek further clarification.

31 Matt Malone, "Lessons from ArriveCAN: Access to Information and Justice During a Glitch," *Intellectual Property Journal* 35, no. 2 (April 2023): 99–139.

Indeed, in the Universal Credit program, welfare agency staff generally do not understand the workings of the automated system that calculates monthly benefit payments. They describe themselves as permanently on the back foot and inadequately trained for the digital service.[32] Crucially, the data that determines auto-generated payments come from employers' reports to the tax office, which precludes staff in the welfare agency from correcting errors.

This opacity and resulting confusion are distinct from the black box dilemma that scholars often scrutinize.[33] Here, algorithmic opacity is dwarfed by the sheer complexity of the spread-out, circuitous bureaucratic system. This design feature means that those who are impacted by administrative decisions, and decision co-producers themselves, cannot easily understand why errors exist nor identify who might be able to remedy problems.

This situation also layers and concentrates administrative burdens because the practical challenges of digital government are distributed unequally and compound with each individual-state interaction. People whose lives are deeply and regularly impacted by bureaucratic institutions are often marginalized: sole-support parents, people with disabilities, people living in poverty, individuals with unsettled immigration status, and members of racialized communities. They may be targeted by state agencies (child protection, for instance), and they may require a privilege or benefit that only a state agency can grant, such as disability benefits or regularized immigration status. Individual administrative burdens may coalesce and multiply, so that these individuals experience the datafied state as oppressive, unpredictable, and impenetrable.[34] Though a "user-friendly" web portal may provide seamless interactions for those for whom the datafied state works relatively well, marginalized individuals seeking to understand or challenge an error may experience the same portal as an incomprehensibly opaque wall.

32 Kayleigh Garthwaite, Jo Ingold, and Mark Monaghan, "Universal Credit and the Perspectives of Ex-Jobcentre Plus Staff," *British Politics and Policy* at LSE (blog), January 15, 2019, https://blogs.lse.ac.uk/politicsandpolicy/ex-jobcentre-plus-staff/.

33 Frank Pasquale, *Black Box Society: The Secret Algorithms that Control Money and Information* (Cambridge: Harvard University Press, 2016); Sandra Wachter, Brent Mittelstadt, and Chris Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology* 31, no. 2 (Spring 2018): 841–888.

34 Wendy Hui Kyong Chun, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge: MIT Press, 2021); see also the accounts of migrants in Petra Molnar, "Territorial and Digital Borders and Migrant Vulnerability Under a Pandemic Crisis," in *Migration and Pandemics: Spaces of Solidarity and Spaces of Exception*, ed. Anna Triandafyllidou (Springer: IMISCOE Research Series, 2022), 45–64.

The diffusion of decision-making authority across and beyond conventional bureaucratic institutions also allows public officials to pass the buck, pointing to the many other actors (managers, technicians, software, databases) that are responsible for an outcome. This practice may be a bureaucratic tradition.[35] Datafied state initiatives, however, allow blame to be shifted further afield to even more responsible others, deflecting officials' own contributions as data providers and co-deciders.

Disrupted Legal Accountability

These practical challenges disrupt legal accountability mechanisms. As an institutional form, bureaucracy uses a hierarchical and traceable accountability structure. Conventional legal tools rely upon this structure, but common accountability mechanisms — internal review procedures, external court challenges, and rights claims — are ill-suited to dispersed forms of public administration.

For example, an important component of bureaucratic accountability is civil servants' express commitment to serve the public interest. This commitment is reinforced when public officials are hired, through internal training, and within office culture. It includes an obligation of responsiveness to members of the public who engage with the bureaucratic agency in question, which centers on listening to individuals' concerns. It aims to get decisions right the first time and to provide review opportunities when things go wrong.[36] From a legal perspective, slower bureaucratic processes are necessary to ensure that officials have time to hear those who will be impacted by their decision, consult with expert colleagues, and make informed decisions that serve the public interest.[37] The bureaucratic practice of internal review, where higher-level officials revisit and amend the decisions of lower-level

35 Matthew S. Hull, *Government of Paper: The Materiality of Bureaucracy in Urban Pakistan* (Berkeley: University of California Press, 2012).

36 Michael Adler, "The Future of Administrative Justice," in *The Oxford Handbook of Administrative Justice*, eds. Marc Hertogh, Richard Kirkham, Robert Thomas, and Joe Tomlinson (New York: Oxford University Press, 2022), 623–646.

37 Amanda Clarke, "Digital Government Units: Origins, Orthodoxy and Critical Considerations for Public Management Theory and Practice," SSRN Scholarly Paper, (July 2017), https://doi.org/10.2139/ssrn.3001188.

colleagues, contributes to this internally-facing accountability mechanism. This administrative justice ideal, and its commitment to getting complex decisions right early on, clashes with a developer ideal that prioritizes speedy decision-making and "accurate" (rather than procedurally fair) results. System designers working in digital government units are not subject to the same accountability techniques. Many serve the public on short-term contracts, which impair an in-office accountability culture. Unlike civil servants, contractors who design, maintain, and repair datafied state tools are not bound by an oath to serve the public good.[38] They are also too far removed from the core of a bureaucratic agency for their decisions to be evaluated through internal review procedures, even if they materially shape a matter that is being internally reviewed.[39] Similar issues arise with the widespread web of data providers and co-deciders that contribute to digital government operations.

Datafied state initiatives also upset legal accountability mechanisms that rely upon external review primarily because they diffuse practical responsibility so widely. External review by courts, specialized tribunals, and other institutions (auditor generals, ombuds offices) is a well-established mechanism for subjecting bureaucracies to legal standards. In this process, an outside body reviews bureaucratic operations to determine whether they meet specific legal standards. For instance, a court might evaluate whether a decision was made by an open-minded decision-maker, whether the decision-maker relied on relevant information, whether the decision-making process raised corruption or substantive injustice issues, and so on.[40] To successfully launch an external review process, however, people must know who to "name." Typically, this requires that individuals name a responsible government ministry or department or a high-level public official, such as a program director. This naming matters because it establishes the scope

38 Robert Thomas, "Investigating Administration and Administrative Law: Research Questions from Immigration Administration," in *Research Agenda for Administrative Law*, ed. Carol Harlow (United Kingdom: Edward Elgar Publishing, 2023), 52.

39 Adelmant and Tomlinson, "Who Builds Digital Government?"

40 Hertogh, Kirkham, Thomas, and Tomlinson, *Oxford Handbook of Administrative Justice* (New York: Oxford University Press, 2021).

of the external review process. It is also fundamentally important to ensure that any remedies ordered at the end of the review (monetary damages, legislative or policy changes) are legally enforceable against an actor who can implement them effectively.

Though decision-making power has long been diffused by bureaucracy, datafied state initiatives disperse decision-making authority much more broadly, making it nearly impossible for members of the public and even lawyers to name appropriate entities. When many actors provide data and co-produce decisions affecting individuals' entitlements and rights, and when technicians translate legal rules into software and databases that shape how decisions are generated, arguably the whole network ought to be reviewed. But both practical opacity and the legal technicalities of external review combined prevent people from easily naming the entities responsible for a particular decision and from holding the operations of those entities to account. The decision-making system itself shields these actors from view, and the legal claiming process often prohibits naming entities situated outside of the core of government. Neither challenge is new, but the practical and legal barriers to accountability can no longer be ignored by anyone concerned with justice in the datafied state.

CORPORATE
CORPORATE
CORPORATE

BY

BURCU BAYKURT

CAPTURE
CAPTURE
CAPTURE

CORPORATE CAPTURE

By Burcu Baykurt

The growing reliance on computational infrastructures in public agencies fundamentally transforms what counts — and is counted — in government. This essay examines corporate capture as a critical feature of the datafied state to demonstrate how the political economy of data-driven technologies shapes statecraft in the digital age.

Corporate capture, in broad strokes, refers to how companies attempt to influence and control governance to advance their interests. This phenomenon, also known as regulatory capture, includes strategies such as lobbying, public relations campaigns, direct contributions, privatization, and policy interference.[1] Many industries use these methods to minimize state intervention, advocate for market-driven policymaking, and shape policy debates, often at the expense of the public interest.[2]

While corporate capture in the datafied state resembles regulatory capture in other sectors, the evolving ties between the state and the tech industry also introduce a range of novel techniques and meanings of corporate capture. In this essay, I discuss two types of capture to explain how the datafied state interfaces with tech companies.

The first type, corporate capture, refers to the ways tech firms exert economic and political influence over the state. I specifically focus on the narratives and modalities of influence taken up by contemporary tech companies to establish and sustain their sway over governments, which increasingly rely on data-driven techniques. The second type, capture corporations,

1 Daniel Carpenter and David A. Moss, *Preventing Regulatory Capture: Special Interest Influence and How to Limit It* (Cambridge: Cambridge University Press, 2013).

2 Anne L. Washington and Joanna Cheung, "Public Interest," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

describes how the industry's capture of the datafied state has created an entirely new market.[3] I highlight the ways tech companies, small and large, build new businesses by capitalizing on captured public data and services, transforming them into new commodities. My empirical analysis focuses on cases originating from Silicon Valley and circulating in a Euro-Atlantic context, but I hope the conceptual discussion invites a conversation with scholars who trace capture practices in the global majority.

As the state undergoes datafication, I suggest that capture by the tech industry extends beyond safeguarding market interests. Companies assume responsibility for delivering public services, adopt state-like roles, and develop commercial ventures harnessing public datasets and services. These novel entanglements raise questions about accountability, equity, and democratic governance. They also fundamentally challenge state capacity, the concept of the public interest, and the prospects of political resistance. This essay aims to inform these critical questions by discussing the shifting practices and scope of corporate capture.

Corporate Capture and the Tech Industry

Tech companies, like their counterparts in other sectors, use lobbying, public relations campaigns, academic research, and industrial action to minimize state intervention and promote market-oriented policymaking. The industry is well-versed in using multiple strategies to shape the terms of policy debates and positioning its interests beyond the reach of regulation.[4] Dating back to the 1990s, Silicon Valley's exponential growth, driven by ubiquitous data collection, has been facilitated by state subsidies and policymakers refraining from interference.[5] Behind the facade of innovation and growth,

3 Thanks to Patrick Davison for coining "capture corporations" to distinguish these two types of capture in the datafied state.

4 Meredith Whittaker, "The Steep Cost of Capture," *Interactions* 28, no. 6 (November 2021): 50–55, https://doi.org/10.1145/3488666; Wendy Y Li, "Regulatory Capture's Third Face of Power," *Socio-Economic Review* 21, no. 2 (April 2023): 1217–45, https://doi.org/10.1093/ser/mwad002.

5 Margaret O'Mara, *The Code: Silicon Valley and the Remaking of America* (New York: Penguin Publishing Group, 2019); Matthew Crain, *Profit over Privacy: How Surveillance Advertising Conquered the Internet* (Minneapolis: U of Minnesota Press, 2021); Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018).

governments have consistently avoided intervening in the emerging tech industry, which has now evolved into the dominant big tech byword.[6]

The captive ties between government agencies and the industry deepened during the post-2008 austerity era. Faced with dwindling public funds and persistent criticisms of inefficiency within the state, policymakers turned to the tech industry's vast capital reserves and extensive computational infrastructure as potential solutions. Through initiatives like smart city programs, civic tech partnerships, or digital inclusion efforts, tech companies have increasingly partnered with local, state, regional, and national governments. Public agencies eagerly embraced the participation of the tech industry to help modernize governments' decaying information systems while testing work-in-progress, novel technologies that could upgrade public services. Tech companies welcomed this invitation to maintain friendly relationships with public agencies, bolster their interests, and seek new avenues of expansion and experimentation.

The industry soon extended its influence into civic and academic initiatives as well, ultimately promoting a new model of government that acts like a platform.[7] Particularly in the 2010s, many civic and academic organizations advocated for investment in open data infrastructures, fostering collaboration with technologists, and adopting a startup mindset within public agencies.[8] Public officials often boasted about fully embracing new technologies and treating government operations akin to entrepreneurial ventures. With data-driven technologies integrating into state capacity, tech companies have started claiming a stake in governance.

While proponents argue that the emerging partnerships between firms and government are win-win, it is apparent that many are established on unequal terms. The tech industry's capital dominance and monopolistic

6 Kean Birch and Kelly Bronson, "Big Tech," *Science as Culture* 31, no. 1 (January 2022): 1–14, https://doi.org/10.1080/09505431.2022.2036118.

7 Tim O'Reilly, "Government as a Platform," *Innovations: Technology, Governance, Globalization* 6, no. 1 (January 2011): 13–40, https://doi.org/10.1162/INOV_a_00056.

8 Burcu Baykurt and Christoph Raetzsch, "What Smartness Does in the Smart City: From Visions to Policy," *Convergence* 26, no. 4 (August 2020): 775–89, https://doi.org/10.1177/1354856520913405.

control over knowledge production underlie its asymmetrical relationship with governments.[9] Tech companies also exploit the weakening of public administration,[10] especially during the post-2008 austerity era, to assert their superiority. By highlighting the inefficiencies of government agencies, they present their prowess in computational infrastructures and data science as the epitome of expertise. This narrative, coupled with ongoing budget cuts and downsizing in public institutions, which set up government agencies to be incapable of delivering on their mandates, aims to infantilize governments and promote privatization. It also seeks to establish tech firms as indispensable partners or even substitutes in governance.

As a result, tech companies increasingly secure government contracts to streamline or deliver public services, while promising substantial cost savings. One example is fraud-detection systems, which use machine learning models as a means to detect fraud. Public agencies adopt these opaque systems without thoroughly reviewing how they make decisions. A few well-publicized scandals have already revealed that fraud-detection systems often falsely accuse people and perpetuate discrimination against minoritized groups.[11] Similar issues arise in risk assessment algorithms in criminal justice, predictive policing, or refugee flow forecasting. Despite high-profile scandals, tech companies and public agencies continue launching and experimenting with these algorithmic systems, often with insufficient public scrutiny.

In addition to being summoned by public agencies, tech companies leverage their perceived economic and epistemological superiority to position themselves as key actors in policy making. Take Google's Environmental Insights Explorer (EIE). In 2018, the tech giant launched a global data platform aimed at helping cities monitor their carbon emissions and develop

9 Amba Kak, Sarah Myers West, and Meredith Whittaker, "Make No Mistake — AI Is Owned by Big Tech," *MIT Technology Review,* December 5, 2023, https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/.

10 Amina Abdu and Abigail Jacobs, "Public Administration," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

11 Morgan Meaker, "The Fraud-Detection Business Has a Dirty Secret," *Wired*, March 7, 2023, https://www.wired.com/story/welfare-fraud-industry/; Gabriel Geiger et al., "Suspicion Machines," Lighthouse Reports, March 6, 2023, https://www.lighthousereports.com/investigation/suspicion-machines/.

climate action strategies.[12] Google promotes the EIE as an opportunity for cities to "access Google's mapping data and ML [machine learning] capabilities" freely, thereby asserting itself as an indispensable partner to governments. However, besides greenwashing Google's substantial contribution to global emissions, the EIE simplifies a complex climate action plan by framing it as a data issue. It tries to portray tech companies as having a supposedly vital role in policymaking as gatekeepers of data capabilities.[13]

In some ways, corporate capture in the datafied state is similar to regulatory capture in other sectors, wherein the industry cozies up to government agencies to safeguard its interests and evade regulation. What sets it apart is the growing transactional ties, where the tech industry deliberately exploits cash-strapped public institutions and asserts its superior expertise in data science. Especially since the 2010s, under the guise of public-private partnerships, tech companies have effectively turned the datafied state into a reliant client.[14] In these so-called partnerships, companies feign a commitment to sharing the risks and responsibilities of modernizing the delivery of public services. However, the terms of the partnerships are rarely equal, and these initiatives do no more than facilitate the tech industry's capture of the state.

In the datafied state, tech companies weave their commercial interests with governments while using public agencies as sites of experimentation for work-in-progress software. Public agencies often shy away from regulating the business model of the industry, have trouble enforcing contracts, and cannot thoroughly review the scope of data-sharing or ownership.[15] Moreover, the industry's solutions essentially transform the intent behind public services. Fraud-detection systems, for example, reframe a social security guarantee for the most vulnerable (welfare benefits) as a task of

12 See: https://insights.sustainability.google/?hl=en-US

13 Eric Nost and Emma Colven, "Earth for AI: A Political Ecology of Data-Driven Climate Initiatives," *Geoforum* 130 (March 2022): 23–34, https://doi.org/10.1016/j.geoforum.2022.01.016.

14 Matthew Bui and Bianca Wylie, "Public-Private Partnership," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

15 Burcu Baykurt, "Algorithmic Accountability in US Cities: Transparency, Impact, and Political Economy," *Big Data & Society* 9, no. 2 (July 2022): 20539517221115426, https://doi.org/10.1177/20539517221115426.; Lilly Irani and Cedric Deslandes Whitney, "Broken Promises of Civic Innovation: Technological, Organizational, Fiscal, and Equity Challenges of GE Current CityIQ," *UC San Diego*, December 23, 2022, https://escholarship.org/uc/item/96q771w6.

optimizing public funds. Ultimately, tech companies become "public actors without public values" in governance.[16] Their interests and techniques take precedence over the common good and fundamentally challenge the notion of the public interest.

A New Kind of Tech Capture: Capture Corporations

As the datafied state grows, corporate capture extends its reach and changes form. I suggest that a new type of capture has emerged in recent years, which I will call "capture corporations." Capture corporations intend to transform the datafied state into a new frontier for the tech industry. Beyond privatization or outsourcing public services, capture corporations seek to build new businesses by commodifying captured government data and services for other industries such as logistics, health care, urban planning, or other governmental or intergovernmental agencies.

I turn to Philip Agre's theorizing of capture to elucidate capture corporations.[17] Though he initially talks about models of privacy, Agre's discussion offers a helpful framework for thinking about data capitalism and state power.[18] Drawing from computing practices, Agre defines capture as the process of restructuring human activities into a computer system's languages. This parsing of human activities, he argues, is not a mere translation but an "active intervention in and reorganization of [human] activities."[19] He also suggests that given capture reduces transaction costs of economic actors, it may usher in a "trajectory toward an increasingly detailed reliance upon (or subjection to) market relations."[20] Building on this, I suggest that

16 Linnet Taylor, "Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector," *Philosophy & Technology* 34, no. 4 (December 2021): 897–922, https://doi.org/10.1007/s13347-020-00441-4.

17 Philip E. Agre, "Surveillance and Capture: Two Models of Privacy," *The Information Society* 10, no. 2 (1994):101–27, https://doi.org/10.1080/01972243.1994.9960162.

18 Alexander Galloway, "Agre > Zuboff," November 10, 2022, http://cultureandcommunication.org/galloway/agre-zuboff.

19 Agre, "Surveillance and Capture," 107.

20 Agre, "Surveillance and Capture," 121.

tech companies' computational capture (i.e., grabbing and parsing) of vast amounts of government data and services over the last few decades has given rise to capture corporations, that is, new practices for market-making.

The shift in the business of computing toward a software as a service model has paved the way for capture corporations. The rise of cloud infrastructures creates new interdependencies between public institutions and big tech companies such as Amazon, Google, and Microsoft.[21] Government agencies become prime targets for expanding the industry's new and often excessive capabilities. Capture corporations also exploit governments' adoption of behavioral approaches in policymaking and their desire to turn public spaces into sensor-driven environments.[22] As a result, the nature of capture in the datafied state evolves from a transactional relationship into an extractive one. Tech giants and startups alike seek to grab more public data and services, repurpose them as new products, stake a claim in public revenues, or bind government agencies to new platforms or subscription services.

One striking example of capture corporations is Amazon's agreement with the UK Health Service (NHS) in 2019, wherein Amazon gained free access to healthcare information collected by the NHS. The deal allowed the company to "create new products, applications, cloud-based services and/or distributed software" and share the information with third parties.[23] Critics rightly pointed out the lack of transparency in the process and the upholding of commercial interests over the public interest.[24] In response to public outcry, NHS officials stressed that no patient data were being shared and the information provided to Amazon was already available online.[25] But this case illustrates a tech giant's rapacious capture of a critical public infrastructure. It exemplifies how public agencies, under the guise of adapting to the

21 Seda Gurses and Joris van Hoboken, "Privacy after the Agile Turn," *SocArXiv*, May 2, 2017, https://doi.org/10.31235/osf.io/9gy73; Taylor, "Public Actors."

22 Marion Fourcade and Jeffrey Gordon, "Learning Like a State: Statecraft in the Digital Age," *Journal of Law and Political Economy* 1, no. 1 (2020), https://doi.org/10.5070/LP61150258.

23 Amy Walker, "NHS Gives Amazon Free Use of Health Data under Alexa Advice Deal," *The Guardian*, December 8, 2019, https://www.theguardian.com/society/2019/dec/08/nhs-gives-amazon-free-use-of-health-data-under-alexa-advice-deal.

24 Taylor, "Public Actors"; "Alexa, What Is Hidden behind Your Contract with the NHS?" *Privacy International*, June 7, 2023, http://privacyinternational.org/node/3298.

25 Elisabeth Mahase, "Government Hands Amazon Free Access to NHS Information," *BMJ* 367 (December 2019): l6901, https://doi.org/10.1136/bmj.l6901.

digital era, also underwrite tech giants' foray into new markets by giving away public information at no cost.

Public institutions also enter into new kinds of revenue-sharing or licensing agreements with capture corporations. The city of Toronto, for example, has partnered with PayIt, a cloud provider of digital payments for governments, to "streamline how residents pay their property taxes, parking tickets, and other municipal services."[26] The deal ensures that PayIt receives a portion of each payment made through the platform. In other words, by becoming an intermediary between the city and residents, PayIt gains a share of the city's public revenues while establishing a lock-in situation on its platform for residents and local government. Similarly, in 2022, the Federal Communications Commission (FCC) in the United States contracted CostQuest to overhaul nationwide broadband maps, which are crucial in distributing federal funds for broadband deployment. Although the FCC and other government entities contributed to the creation of the CostQuest database, the final product, the FCC National Broadband Map, is considered proprietary. Access to the map is only possible through a licensing fee for public and private institutions.[27] Both of these cases exemplify how capture corporations increasingly seize public information and turn it into proprietary products while fostering dependencies for government agencies.

The growth of capture corporations has ultimately spurred a new market known as GovTech. [28] Several small to midsize startups now compete to transform public services and data into new business ventures. These firms specialize in various areas, such as optimizing utility management, regulating curbs and parking spaces, and providing data analytics services to government agencies and private companies. GovTech, still a nascent and somewhat ambiguous market, illustrates the ambition of the industry to

26 Samantha Beattie, "Toronto Council Approves $20M Deal with U.S. Tech Company PayIt, but Some Competitors Crying Foul," *CBC*, May 5, 2021, https://www.cbc.ca/news/canada/toronto/toronto-council-approves-20m-deal-with-u-s-tech-company-payit-but-some-competitors-crying-foul-1.6015431.

27 Christopher Ali, "Where the Market Dares Not Tread: Mapping Rural Broadband in the United States," in: *Media Rurality*, eds. Patrick Brodie and Darrin Barney (Durham: Duke University Press, forthcoming); David B. McGarry, "Panelists at Broadband Breakfast Event Urge the FCC Mapping Fabric Be Made Public," *Broadband Breakfast*, September 23, 2022, https://broadbandbreakfast.com/2022/09/panelists-at-broadband-breakfast-event-urge-the-fcc-mapping-fabric-be-made-public/.

28 Nitesh Bharosa, "The Rise of GovTech: Trojan Horse or Blessing in Disguise? A Research Agenda," *Government Information Quarterly* 39, no. 3 (July 2022): 101692, https://doi.org/10.1016/j.giq.2022.101692.

capitalize on the datafied state as a profitable domain. Particularly in the United States, it exploits the image of a slow and cumbersome government bureaucracy and claims legitimacy via the widely accepted practice of contracting public services out to private companies in the name of efficiency.[29] But instead of selling software and data to the public sector, these firms extract value from government services and data, creating new avenues for profit. GovTech deeply intertwines the datafied state and the tech industry, surpassing the realm of regulatory capture.

It is difficult to anticipate whether these changes will strengthen or weaken state capabilities and what kind of counteract measures public agencies and counterpublics[30] may develop against capture corporations. Thus far, it appears public officials rarely consider the consequences of expanding capture corporations beyond their initial purpose, potentially becoming intermediaries for a broader range of transactions and interactions.[31] There is not enough discussion about whether GovTech firms — and the profitability of government services and datasets — might dictate how public officials define or prioritize social problems. Nor is there a conversation about where to draw the line when it comes to embedding these firms in public governance.[32] In addition to undermining accountability and corrupting public agencies, capture corporations may compromise civic capacity too. The abundance and accessibility of data,[33] driven by their lucrative prospects, might shape the trajectory of data publics.[34]

Conclusion

Corporate capture in the datafied state comes in many forms and degrees. This essay intended to offer conceptual clarity on the shifting practices and extent of capture. I have discussed that tech companies solidify their

29 Jennifer Raso and Victoria Adelmant, "Bureaucracy," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024); Since the Federal Activities Inventory Reform (FAIR) Act of 1998, US federal government agencies have been pushed to carry out their functions by procuring from and partnering with private companies. See: https://www.govexec.com/federal-news/1998/10/clinton-signs-privatization-bill/4725/.

30 Seyi Olojo, "Counterdata," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024); Vanessa Massaro, Darakhshan J. Mir, Terrell Mosley, and Nathan C. Ryan "Counterdata," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

31 Matthew Claudel and Bianca Wylie, "Technology Procurement: Shaping Future Public Value," *Open North*, February 1, 2021, https://opennorth.ca/resources/technology-procurement-shaping-future-public-value/.

32 Hannah Bloch-Wehba, "A Public Technology Option" *Law & Contemporary Problems* 86 No. 3, (forthcoming): 23-41, https://ssrn.com/abstract=4509051.

33 Fleur Johns, "Governance by Data," *Annual Review of Law and Social Science* 17 no. 1, (October 2021): 53–71, https://doi.org/10.1146/annurev-lawsocsci-120920-085138.

34 Youngrim Kim, "Data Publics," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

influence over the state through increased lobbying, seizure of public resources, and emphasis on computational competence. Government agencies outsource public services to tech firms, implement automated administrative tools, and enter data- and revenue-sharing agreements. The tech industry further infiltrates the state by creating new dependencies via cloud computing infrastructures and data analytics tools.

This intensifying dependence on technology companies and digital infrastructures indeed undermines the regulatory capacity of public agencies, impeding their ability to regulate the entrenched power of the industry and hold tech companies accountable. As data capitalism and state capacity continue to blend, often seamlessly, we need more on-the-ground documentation of how these captive ties reshape state capacity, perpetuate harms and social stratification, and obstruct the public interest.[35] The techniques and implications of corporate capture may vary in different policy areas (e.g., tax, privacy, or algorithmic accountability); across authoritarian and democratic datafied states; and at local, national, or international levels of policymaking. However, what remains constant is the pressing need to fight against this increasingly concentrated power of the tech industry, demand robust regulatory action from governments, and radically rethink digital infrastructures, especially within the datafied state, in a way that centers the interests of publics.

35 The Institute for Technology in the Public Interest, "Infrables: The Cloud is not an Option," (May 2022), https://titipi.org/pub/Infrables.pdf.

BUREAUCRACY
BUREAUCRACY
BUREAUCRACY

BY

LUDMILA COSTHEK ABÍLIO and CAROLINA CRUZ

BUREAUCRACY
BUREAUCRACY
BUREAUCRACY

BUREAUCRACY

By Ludmila Costhek Abílio and Carolina Cruz

Bureaucracy refers to the systems of people, documents, and regulations that organize the day-to-day operations of the state. Bureaucracy is the basis for the functioning of the state in all its spheres, including public policy management, the legal system, the social security system, among others. Bureaucracy materializes logics, rationalities, decision-making processes, and modes of operation of the state. But what happens to bureaucracy as the state becomes datafied? We argue that the history of bureaucracy and its association with a fetishism of neutrality makes it uniquely vulnerable to arguments that it should be mechanized, automated, and datafied. After all, much of contemporary data technology is marketed on its supposed inherent values of efficiency and neutrality. We use the case of Brazilian President Jair Bolsonaro's 2019 efforts to digitize aspects of the government to demonstrate how the datafication of bureaucracy is inherently political, despite presenting itself as a technical process. We also discuss how datafication processes driven by companies today are redefining the role of state bureaucracy in regulating services and building trust in the private sector.

The management and functioning of the state are carried out through bureaucracy. It is through the bureaucratic apparatus that the state exercises, for example, the monopoly to confer legal existence to individuals. It is also within the state bureaucratic apparatus that the design, implementation, and execution of public policies and the granting of social benefits take place.

German sociologist Max Weber wrote about the role of rationalization in modern capitalist society to theorize bureaucracy. Rationalization can be recognized in bureaucratic state administration, in labor management and organization, in the field of science, and in the conduct of life. Modern bureaucracy is based on impersonality, efficiency, and calculation. Its governing body is made up of employees guided not by values or interests, but by impersonal and legally-established laws, rules, and procedures, which are focused on the technical and efficient operation of public administration.[1] In the early 20th century, Weber had already pointed to the irrationality that permeates this rationalization, in a bureaucratic apparatus that operates in a dehumanized, automated way, focused on practical purposes.[2]

Underlying bureaucratic rationalization is what we're calling a fetishization of neutrality. That is, state bureaucracy is seen as valid, fair even, when it operates according to consistent, inflexible, and legally-determined procedures.[3] Thus, rationalization as a means of achieving neutrality is seen as politically valuable, even as it de-humanizes the process. This is the irrational rationalization that Weber describes, and is the contradiction that collides with contemporary attempts at datafication.

Dafication has long been imagined as a vehicle for promoting efficiency and objectivity, operating through purely technical means that surpass human capabilities.[4] However, the processes of digitization in bureaucracy involves a perverse encounter that enhances the fetish of neutrality, both of technology and bureaucracy. The datafication of bureaucracy incorporates, in an obscure and challenging-to-map manner, inequalities and social dynamics into selection, methods, and parameters for verifying citizens' eligibility, among other elements. Consequently, the datafication of bureaucracy can produce or reproduce, in an unpredictable or unplanned way, social inequalities and injustices.[5]

1 Max Weber, *Economy and Society.* Vol.01, (Univ. of California Press, 1978).

2 Max Weber, *The Protestant Ethic and The Spirit of Capitalism*. The revised 1920 edition (New York and Oxford: Oxford University Press, 2011).

3 Amina Abdu, Abigail Jacobs, "Public Administration," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

4 Tarleton Gillespie, "The Relevance of Algorithms," in *Media Technologies: Essays on Communication, Materiality and Society*, eds. Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot (Cambridge, MA: MIT Press, 2014) 167–193.

5 Virginia Eubanks, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018); Cathy O´Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishing Group, 2016); Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

In Brazil, the government of Jair Bolsonaro adopted a series of initiatives aimed at accelerating digitization in public services. Measures were implemented to promote the integration of various databases, as well as mechanisms for centralizing population data management. The government's description and justification for these efforts clearly demonstrates the fetish of neutrality that runs between bureaucracy and data technologies. Such measures were defended as inherently apolitical, but masked the ways in which the supposed reduction of bureaucracy actually extended its power.

In 2019, the government created the Special Secretariat for Debureaucratization, Public Administration and Digital Government (Secretaria da Desburocratização, Gestão e Governo Digital), and in 2020, a strategic plan for digital government. In the name of streamlining bureaucracy, the strategy relied on two fronts: expanding the digitization of public services and integrating databases from different government entities. In 2019, the Citizen Base Registry (CBC) was created by a federal decree. This major database aims to integrate multiple datasets from different government entities and facilitate their access to this centrally maintained one, which includes citizens' biographical, biometric, and registration data. The decree sets objectives such as:

> simplifying the provision of public services, guiding and optimizing the formulation, implementation, evaluation, and monitoring of public policies, enabling the analysis of access conditions and maintenance of social and fiscal benefits, promoting the improvement of the quality and reliability of data held by the federal public administration, and increasing the quality and efficiency of internal operations of the federal public administration.[6]

Based on the defense of streamlining bureaucracy, the government facilitates both integration and flow between various databases. The decree

6 Presidência Geral da República, "Decreto 10.046/2019," Brasília, 2019.

also established a Central Data Governance Committee (Comitê Central de Governança de Dados), which is responsible for defining the guidelines that determine the levels of data restriction on the datasets (broad, restricted, or specific). Initially, this committee consisted only of representatives from government institutions. Responding to demands from the Supreme Court, two representatives from civil society entities were included in the Central Data Governance Committee, and the CBC became subject to the General Data Protection Law in force since 2020. The fact is that Brazil is building a centralized and fluid state database, implemented without involvement or debate with civil society entities. The use of such data infrastructure to bypass public oversight is far from a neutral, bureaucratic act.

Of further concern is that while pushing to dataify the state, the Bolsonaro government signaled its intentions to both facilitate private entities' access to state databases, as well as integrate data management, surveillance, and state security policies.[7] The result would be a new means of citizen data flowing to other actors. Among the federal entities that had already requested access to the CBC in 2019 were the Army Command and the Brazilian Intelligence Agency.[8]

With the election of Luis Inácio Lula da Silva, the datafication of bureaucracy and public services continue. It will be necessary to investigate over time the regulatory mechanisms and paths taken by this administration in the extraction and uses of citizens' data.

The fetish of neutrality also obfuscates the role of corporations that oligopolize the means of digitization. These corporations present themselves as intermediaries — (neutral) providers of technology in various fields. And yet, mapping, explaining, or delimiting the power of these corporations is incredibly difficult.

7 Coding Rights, *Cadastro Base do Cidadão: A Megas de Dados* (Rio de Janeiro: Coding Rights, 2020).

8 Coding Rights, *Cadastro*.

While data is being used to streamline and control certain aspects of governance, such as national security and taxation, it has also been used to relinquish government oversight in other aspects, such as the management of labor.[9] We refer to this as the uberization of work that manifests along three interrelated social transformations: the informalization of bureaucratic practices that govern labor processes; the transformation of workers into just-in-time workers; and the centralization of labor control through oligopolies.[10] Again, data is central to these processes, as private companies such as Uber take over the organization of certain aspects of public life (public transportation, for example) by espousing neutrality in organizing the relationship between riders and drivers on public roads.

With the uberization of work, there is also a noticeable shift away from the centrality of the state in conferring trust and legality to services offered by the private sector. Historically, state bureaucracy has been the locus for processing and enforcing certifications and regulations that grant legality to the private sector. Companies like Uber construct trust and certification for their services through new means, challenging the power of the state. In this way the datafication processes of work management serve as mechanisms to confer legitimacy, no longer relying on legally-established regulations but on the trust built through the actions of a multitude of vigilant consumers who monitor and certify the quality of work.[11]

The datafication of the bureaucratic apparatus is based on the defense of efficiency, technique, and impersonality within the state; however, it produces modes of control and surveillance over citizens as well as transfers state control to private entities. Both these developments challenge emerging regulations around data protection in countries such as Brazil, which must navigate competing private interests and sociopolitical uses of data.

9 Burcu Baykurt, "Corporate Capture," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

10 Ludmila Abilio, "Uberization: The Periphery as the Future of Work?" in *Platformization and Informality*, eds. Aditi Surie and Ursula Huws, (Switzerland: Palgrave MacMillan, 2023).

11 Abilio, "Uberization."

DATA
DATA
DATA

BY

YOUNGRIM KIM

PUBLICS
PUBLICS
PUBLICS

DATA PUBLICS

By Youngrim Kim

Governments today are undergoing a digital transformation, actively designing, developing, and implementing computational tools and algorithms to improve the efficiency of public administration and services. This shift toward the datafied state entails the laborious task of converting vast amounts of government and public sector information into machine-readable formats. In democratic regimes, the disclosure of this extensive public sector data and making it accessible for reuse have become critical benchmarks for assessing government transparency and accountability.

This process of infrastructuring public datasets engages many old and new actors, from people who build and maintain "public" data infrastructures to those who monitor or repurpose these newly available datasets. *"Data publics" refer to these heterogeneous groups of people who build, maintain, and use public data infrastructures as a means of civic engagement.* Consequently, the concept of data publics raises important questions regarding the politics of civic engagement and participatory governance within the datafied state. Studies on data publics have explored who is capable of and encouraged to participate in this new mode of civic engagement, as well as the political potentials and limitations of data publics.[1]

Based on my ethnographic research of Korean open data communities during the COVID-19 pandemic, I will discuss how different formations of data publics evolve in relation to local open data initiatives. South Korea — a postcolonial, post-authoritarian country with a history of rapid and tumultuous democratization — serves as a valuable site to examine the conceptual

1 Evelyn Ruppert, "Doing the Transparent State: Open Government Data as Performance Indicators," in *The World of Indicators*, eds. Richard Rottenburg, Sally E. Merry, Sung-Joon Park, Johanna Mugler, and Evelyn Ruppert (Cambridge: Cambridge University Press, 2015), 127–150; Anne L. Washington, "Who Do You Think We Are? The Data Publics in Digital Government Policy," in the *Proceedings of the 52nd Hawaii International Conference on System Sciences* (January 2019): 3264–3272; Jean Burgess, Kath Albury, Anthony McCosker, and Rowan Wilken, "Everyday Data Publics," in *Everyday Data Cultures* (Cambridge: Polity, 2022), 115–143.

limitations of existing models of data publics. While the concept of data publics is rooted in the western notion of publics as a counterweight to state practices, the Korean case illustrates how data publics can complement, and even partner with the state, particularly when framed within the affective relations of patriotism.

The purpose of this essay is to highlight the importance of connecting contextualized and historically grounded accounts of public formation and evolution to the study of data publics. In the current environment of the datafied state, how have these groups evolved and who have emerged as new significant stakeholders? To address these questions, the essay brings together scholarship from Asian and Korean cultural studies that challenge the western-centric theorization of the "public." By illustrating how local configurations of the state, market, and civil society have shaped a different formation and functioning of publics in South Korea, I urge the need to re-formulate the concept of data publics to incorporate these historically driven, local manifestations of the public. Only then can we decenter the study of global data cultures without relegating those outside the Anglo-American world to sites of difference.

Global Open Data Movement and the Emergence of Data Publics

The emergence of the global open data movement in the early 2010s aimed to enhance government transparency and accountability by releasing public sector information in digital formats. Open data initiatives were strongly influenced by the open-source movement in the 1990s–2000s, which lies at

the heart of the Silicon Valley ideology of free software and counter-culture libertarianism.[2] In the United States, open data became one of the central pillars of President Obama's Open Government Initiative in 2009, with the goal of making government information machine-readable for all. By making these datasets available, governments committed to ensuring transparency and accountability to their citizens, while also expecting to drive innovation in public services and foster new businesses through data reuse. Collectively, the global open government data movement was based on these three key foundations[3]:

- *Transparency:* Enabling citizens to monitor government activities and initiatives
- *Social and commercial value creation:* Promoting opportunities for innovation and commercialization through the release of public sector data
- *Participatory governance:* Empowering citizens to actively engage in public decision-making and policy development

International organizations like the United Nations and the Organisation for Economic Co-operation and Development (OECD) have heavily promoted open government data (OGD) to their members and partner countries as an indicator of mature and innovative democracy. OECD, in particular, created the Open-Useful-Reusable data index (OURdata) to assess government efforts to support OGD.

Since the enactment of the Electronic Government Act in 2001, the South Korean government has pursued digital government as a core policy for national development. Particularly, OGD materialized through President Roh Moo-hyun's e-government initiatives in the early 2000s and the subsequent

2 Simon Chignard, "A Brief History of Open Data," *ParisTech Review*, March 29, 2013, https://www.paristechreview.com/2013/03/29/brief-history-open-data/.

3 Judie Attard, Fabrizio Orlandi, Simon Scerri, and Sören Auer, "A Systematic Review of Open Government Data Initiatives," *Government Information Quarterly* 32, no. 4 (October 1, 2015): 399–418, https://doi.org/10.1016/j.giq.2015.07.006.

Park Geun-Hye administration's Government 3.0 — a master plan for a new governance paradigm that puts forth openness, sharing, communication, and collaboration with civic actors as founding principles of public sector reform.[4] As part of Government 3.0, President Park signed the Public Data Act in 2013, which mandated the disclosure of government and public sector data in machine-readable formats. Since then, South Korea has consistently ranked first in OECD's OURdata index for three consecutive years (2017, 2018, 2019), receiving high scores in data availability, data accessibility, and government support for data reuse.[5]

These top metrics in open data initiatives meant more than mere statistics in South Korea — they were celebrated as symbols of national achievement and international recognition of South Korea's digital prowess. The Korean state actively promoted these successes as evidence that "the world is recognizing the Korean government's digital competitiveness" and as a demonstration of South Korea's leadership in driving global digital government transformation.[6] Promoting digital government as a national project is a continuation of South Korea's history of techno-nationalism. Harnessing sociotechnical imaginaries of Korea's digital infrastructure projects as symbols of modernity encapsulates the developmentalist desire to overcome the national traumas of war and colonialism through technological advancements. In other words, the Korean state viewed open government initiatives as nation-building opportunities to showcase the country's global competitiveness in technology, digital innovation, and democratic infrastructures.

Within these global and localized contexts of the open data movement, the term "data publics" emerged to describe new groups of people who responded to OGD across the globe. As the objective of OGD was not just to establish a technical foundation but to encourage citizens to participate and

4 B. Shine Cho and Sangoh Yun, "Yŏllin chŏngbu shidaeŭi shiminch'amyŏ: shibik'aek'ingŭi yuhyŏng pullyu yŏn'gu [Citizen Participation for Open Government: A Typology of Civic Hacking]," *Han'gukchŏngch'aek'ak'oebo* [*The Korean Association of Policy Studies*] 26, no. 1 (2017): 177–202.

5 OURdata index evaluates the government's effort to support open government data through three criteria: Data availability, which measures "the extent to which governments have adopted and implemented formal requirements to promote open government data at the federal/central level"; data accessibility, which measures "the extent to which government data are provided in open and re-usable formats"; and lastly, government support for data reuse, which measures "the extent to which government play a proactive role in promoting the re-use of government data inside and outside the government" (see Guillaume Lafortune and Barbara Ubaldi, "OECD 2017 OURdata Index: Methodology and Results," *OECD Working Papers on Public Governance*, no. 30 (December 11, 2018): 1–45, https://doi.org/10.1787/2807d3c8-en).

6 Ministry of Interior and Safety (MoIS), "South Korea Tops the 1st OECD Digital Government Evaluation: Amid Crisis, the World Recognizes South Korea's Digital Government Competitiveness," October 16, 2020, https://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=80502.

collaborate in governance processes, it mobilized a multiplication of publics who would monitor and reuse OGD to "actively witness the affairs of the state."[7] According to Evelyn Ruppert, data publics are not mere recipients of data; instead, they "are incited to do their own experiments, establish matters of fact, see the state for themselves and disseminate their results to others … data publics are constituted by dynamic, complex, and uncertain arrangements of actors mobilized and provoked by open data."[8] These people include civic hackers, data journalists, and activists who convene through data portals, Freedom of Information Act requests, and other platforms to reuse OGD according to their interests and capabilities.

In this sense, the concept of data publics is heavily influenced by western political notions of the "public"[9] and the "public sphere,"[10] which emphasize the significance of publics in fostering critical public deliberation free from state intervention and economic pressures. Similarly, in existing scholarship, data publics are envisioned as critical civic actors in today's democratic systems. They are expected to enhance civic engagement by monitoring and utilizing public sector data. This western conception of data publics is premised on these civic actors acting as a counterbalance to state practices, those who can foster independent civil society using the newly available "tools" (or public sector data).

Cultural studies critiques that scrutinize the idealized portrayal of a universal public sphere and its exclusionary tendencies have been useful in identifying the politics of inclusion within the realm of data publics.[11] Who is invited or granted access to different formations of data publics? For instance, Anne L. Washington sharply points out that open data initiatives have promoted participation and collaboration between the government and civic actors without a clear delineation of who constitutes this public.[12] Open government policies, therefore, have exhibited limitations

7 Washington, "Who Do You Think You Are?" 3266.

8 Ruppert, "Doing the Transparent State," 3.

9 John Dewey, *The Public and Its Problems* (Athens: Swallow Press, 1927).

10 Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (Cambridge: MIT Press, 1991).

11 Nancy Fraser, "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy," *Social Text*, no. 25/26 (1990): 56–80, https://doi.org/10.2307/466240.

12 Washington, "Who Do You Think We Are?"

in representing the diverse consumers of open data and in recognizing the varying technical capabilities possessed by these consumers.

Furthermore, despite the rich body of cultural studies scholarship from the Global South that has extensively challenged and reformulated the western notion of "publics," these critiques have not been fully integrated into the current formulations of data publics. In the following section, I will illustrate how critiques from Asian cultural studies can provide valuable insights for efforts to dewesternize the liberal understanding of data publics.

Lessons from Asian Cultural Studies to the Study of Data Publics

Historical and contextual understanding of how publics emerge in non-Western contexts needs to be critically reflected in the formulation of data publics. Asian cultural studies, in particular, contribute to destabilizing the state-society premise that underlies the liberal conception of data publics.[13] The global popularity of the public sphere theory comes with an imagination that assumes civil society resides outside of the state and holds the potential to question, or even overthrow, state power. The rise of public sphere theory in South Korea in the late 1980s was rooted in this hope for *simin* (citizens) to act as agents of political reform and social movements.[14] As South Korea was undergoing a critical transformation from three decades of authoritarianism to a national democratization movement, Korean scholars found the promise of the public sphere theory — a civil society independent from state power and fostering an alliance of enlightened middle-class citizens — extremely appealing. As Jiyeon Kang diagnoses, since Jürgen Habermas' 1996 lecture at the Seoul National University and the following publication of his translated works, the public sphere theory

13 Myung-koo Kang, "Hunmin'gongnonjangŭi ironjŏk kusŏngŭl wihayŏ: habŏmasŭ pilligi, pik'yŏgagi, nŏmŏsŏgi [Toward a Formation of Hunmin Public Sphere: Appropriating and Reformulating Habermas]," *K'ŏmyunik'eisyŏn iron* [Communication Theory] 9, no. 2 (2013): 10–51; Myung-koo Kang, *Hunmin'gwa Kyemong: Han'guk Hunmin'gongnonjangŭi Yŏksajŏk Hyŏngsŏng* [*Didactics and Enlightenment: The Historical Formation of the Korean Didactic Public Sphere*] (Seoul: Nanam, 2016); Jiyeon Kang, "Old and New Questions for the Public Sphere: Historicizing Its Theoretical Relevance in Post–Cold War South Korea," *Media, Culture & Society* 43, no. 1 (January 1, 2021): 158–70, https://doi.org/10.1177/0163443720939480; Kuan-Hsing Chen, "Civil Society and Min-Jian: On Political Society and Popular Democracy," *Cultural Studies* 17, no. 6 (November 1, 2003): 877–96, https://doi.org/10.1080/0950238032000150075.

14 Kang, "Old and New Questions."

has been widely embraced as one of the canons in Korean communication and cultural studies.[15]

However, starting in the 2000s, there has been a growing recognition of Eurocentrism in Asian communication and cultural studies internally.[16] These scholars began to critique the colonial paradigm of global knowledge production — which positions the West as the primary site of theory production while the rest is relegated to case studies for testing these theories — and strived to recover the political capacity of local theories. Myung-koo Kang's reformulation of the public sphere came out in response to these concerns.[17] To dewesternize the Habermasian idea of the public sphere, Kang questioned whether a civil society that is independent and autonomous from the state has ever existed in South Korea. Tracing the genealogy of the Korean public sphere from the Choson Dynasty (1392–1897), he reveals it was an exclusive domain reserved for discussions among the king and his bureaucrats, almost entirely composed of the ruling class men. As reflected in the concept of kong (the public; 공; 公), which means both the "public interest" and the "virtue of the ruler who leads the people," the public sphere at this period was essentially didactic and aimed at indoctrinating the populace to serve the royal dynasty.[18] According to Kang, this didactic nature persisted throughout South Korea's colonial and postcolonial eras with journalists, reformists, and intellectuals serving as pivotal figures in the Patriotic Enlightenment Movement (1905–1910) to subsequent phases of rapid modernization and democratization.[19] As the public sphere led by these power elites aimed to educate other members of the society "under the umbrella of the nation-state, rather than serving the welfare of the people," this period of violent and rapid growth delimited Korea's opportunity to develop and mature an independent civil society.[20] Therefore, patriotism continues to play a vital role in shaping the function of the Korean public sphere.

15 Kang, "Old and New Questions."

16 Tae-Il Yoon, "Han'guk k'ŏmyunik'eisyŏn yŏn'guesŏ sŏgujungshimjuŭi nŏmŏsŏgi [Beyond Eurocentrism in Korean Communication Studies]," Han'gugŏllonhakpo [Korean Journal of Journalism and Communication Studies] 58, no. 6 (2014): 445-474.

17 Kang, "Hunmin'gongnonjangŭi ironjŏk kusŏngŭl wihayŏ [Toward a Formulation of Hunmin Public Sphere]"; Kang, Hunmin'gwa Kyemong [Didactics and Enlightenment].

18 Kang, "Old and New Questions."

19 A nationalist project that was held by Korean intellectuals to reclaim sovereignty from Japanese imperialism. It emerged as a response to the signing of the "Eulsa Treaty" in November 1905, which was forcibly concluded by Japan to deprive Korea of its diplomatic rights. The purpose of this movement was to cultivate the modern capabilities of the Korean people through cultural activities such as education, media, religion, and industrial development. Progressive intellectuals and urban citizens primarily led this movement.

20 Kang, "Hunmin'gongnonjangŭi ironjŏk kusŏngŭl wihayŏ [Toward a Formulation of Hunmin Public Sphere]," 51.

I argue that it is difficult to detach this looming presence of nationalistic aspiration, rooted in Korea's fractured and externally coerced history of modernization, from the formation of data publics in South Korea today. Public and academic discourse on Korea's open data communities and civic tech frame them as reliable partners of the state that could assist in addressing social problems utilizing public data and their advanced digital capabilities. Much of the literature on citizens' engagement with public sector data in Korea — also frequently described with terms like "civic tech" and "civic hacking" — characterizes it as an entrepreneurial mode of citizen-led, public service innovation that shows the potential for citizens to become collaborative partners of the government.[21] Thus, open data communities in South Korea align more closely with the role of "government service developers," who contribute to improving the design and delivery of public services. This attitude has been particularly salient in various open data communities that emerged during the COVID-19 pandemic. The story that follows provides a snapshot of the evolving relationship between the Korean state and data publics, particularly in the context of COVID-19.

Data Publics in South Korea's Digital Response Against COVID-19

In July 2020, the South Korean government implemented nationwide QR code entry log systems in response to the COVID-19 pandemic. This mandate required all individuals to scan their personal QR codes when entering facilities with a high risk of virus transmission, including various public spaces such as bars, cafés, restaurants, libraries, and more. The Korean government urged people to contribute to the collective effort of monitoring COVID-19

21 Cho and Yun, "Citizen Participation"; Jiwoo Hyun and Sangoh Yun, "Saeroun shiminch'amyŏ: saengt'aegye kwanjŏmesŏŭi shibikt'ek'ŭ pigyobunsŏk [New Citizen Participation: Comparative Analysis of Civic Tech from an Ecosystem Perspective]," *Han'gukkonggonggwallihakpo* [*Korean Public Management Review*] 32, no. 3 (2018): 349-379; Joonhyeog Park and Suyoung Kim, "Tijit'ŏl konggonghaengjŏngesŏ shiminch'amyŏ tan'gyee kwanhan yŏn'gu [Citizen Participation in Digital Public Administration]," *Han'guksahoebokchihaengjŏnghak* [*Journal of Korean Social Welfare Administration*] 23, no. 1 (2021):175-205.

by scanning their QR codes and leaving their entry logs in the government's contact tracing database. This data collection aimed to provide health officials and contact tracers with individual location data so they could use the information to identify COVID-19 hotspots. The establishment of this large-scale public health surveillance infrastructure was not solely the result of government and state authorities like the Korea Centers for Disease Control and Prevention and the Central Disease Control Headquarters. Instead, it emerged from an extensive collaboration between government institutions, major Korean tech companies like Naver and Kakao, as well as various civic data communities that played a crucial role in collecting and maintaining COVID-19 data. Cultivating close ties with relevant state officials, these groups created various COVID-19 related digital services reusing public data to assist the government in need. Particularly, civic data communities, coalescing under the name COVID-19 Joint Response Team, offered policy recommendations in open COVID-19 data formats, developed mobile applications that would assist efficient allocation of protective equipment, and formulated privacy-protecting safety codes for the nationwide distribution of the QR check-in system.

A patriotic attitude served as a crucial motivation, as demonstrated by many members of the COVID-19 Joint Response Team with whom I engaged. For instance, one member who participated in developing privacy-protecting safety codes for the government's COVID-19 surveillance infrastructure explained in a public interview that they joined the project to "donate my skill when the country is in need."[22] Despite the government's offer to compensate their work as an outsourcing arrangement, the team declined because they did not want the paperwork to take up too much time. "We wanted to solve this national crisis with the government," another interviewee commented. As reliable partners of the state, they played a vital role in the development of Korea's technocratic, data-driven response to

22 The interviews can be found here (written in Korean): https://codefor.kr/g/home/news/5/13

the pandemic. Here, "pandemic data publics" came into being — those that participated in producing, distributing, and maintaining COVID-19 data infrastructures. These communities established close ties with state institutions and became integrated into the fabric of state governance. Alongside national achievements in e-governance and open data initiatives, these "innovative" forms of civic participation were presented as evidence of Korean citizens' advanced digital capabilities and their democratic usage. In other words, these data publics in South Korea have been heavily co-opted, and absorbed into the Korean state's developmentalist and nation-building project of "Digital Korea."

When the state-society relationship is forged through such affective relations of patriotism, data publics' potential to critique or problematize state affairs becomes limited. Rather than being vigilant observers of the state, they become benign forms of civic participation — participation that the state finds comfortable promoting and incorporating into state governance. Reformulating the concept of data publics allows it to capture such forging of alliances and strategic co-optation between the state and civic data communities. As discussed above, Korean cultural studies on the evolution of the public sphere in South Korea offer a valuable resource to explain these changing dynamics. The affective alignments between the state and data publics must be situated within Korea's local configuration of state-society relationships, which is heavily shaped by nationalistic desires for technocratic futures.

Reformulating Data Publics

The project of dewesternization ultimately lies in repositioning nontraditional sites of knowledge production within the domain of theory work, where new theories are generated and old ones are critically reevaluated. The South Korean context contributes to the concept of data publics by questioning its implicit assumption that data publics exist as separate entities outside of the state. When these data publics' techno-optimistic vision deeply aligns with the state[23] and maintains strong partnerships in materializing these objectives, it becomes challenging to differentiate these actors from institutional forms of governance. Thus, the COVID-19 Joint Response Team illustrates a case where citizens' engagement with public sector data becomes a hegemonic mode of participation mobilized in the name of patriotic duty.[24] To better capture these cases, I suggest reformulating data publics as *a changing configuration of actors involved in developing, maintaining, and using public data infrastructures*, to unsettle its previous assumption of the state/society divide. By focusing on the evolving configurations (and reconfigurations) of heterogeneous actors, this formulation allows for examining the unequal social relations and power dynamics that data publics remediate. For example, this perspective explains how bureaucrats, state authorities, industry players, and civic data communities came into compromised alignments and ambivalent collaborations in cultivating Korea's pandemic data governance. Hence, data publics as an analytical framework is inherently precarious and organic: the question then becomes *when is a data public* instead of *who or what is a data public*.[25]

This reconceptualization of data publics echoes Jonathan Gray's suggestion to view open data as "infrastructural devices", rather than as simply a

23 Avle et al. (2020) explain how techno-optimism becomes a technique of governance that is harnessed for nation-building, particularly in the Global South. Using case studies from China, Ghana, and Indonesia, they demonstrate how "scale" is used as a logic to align different visions and hopes for national futures possessed by state actors and citizens. See Seyram Avle, Cindy Lin, Jean Hardy, and Silvia Lindtner, "Scaling Techno-Optimistic Visions," *Engaging Science, Technology, and Society* 6 (May 2020): 237–254, https://doi.org/10.17351/ests2020.283

24 As a caveat, this is not an attempt to universalize data publics in South Korea. There are indeed multiple data publics as well as counterpublics that engage with public sector data in non-traditional ways, as demonstrated by the examples of Korean feminist organizing (i.e., School MeToo, femiwiki) and urban queer movements (i.e., Seoul Queer Collective). The purpose of this essay is to highlight the importance of connecting contextualized and historically grounded accounts of public formation and evolution to the study of data publics, to understand why certain formations of data publics have become hegemonic in South Korea.

25 Here I echo Star and Ruhleder (2016), who used the phrase, "when is an infrastructure?" (instead of "what is an infrastructure") to highlight infrastructures' relationality. See Susan Leigh Star and Karen Ruhleder, "Steps Toward an Ecology of Infrastrcuture: Design and Access for Large Information Spaces," in Boundary Objects and Beyond, eds. Geoffrey C Bowker, Stefan Timmermans, Adele E. Clarke, and Ellen Balka (Cambridge: MIT Press, 2015), 377–415.

representational resource that frames the politics of data only through the issues of access.[26] Instead, the infrastructure-oriented perspective takes into account the performative capacities of data infrastructures. It is a broader look at how different forms of participation and collaboration become materially organized in the building of data infrastructures. Hence, this reformulation of data publics serves as a useful analytical tool to understand how individuals are assembled through heterogeneous arrangements mobilized by open government and public sector data. It is a relational analytic — focusing on the associations and dissociations of various groups of people that engage in activities ranging from building and maintaining data infrastructures to normalizing governments' data practices to problematizing and resisting them.

This essay emphasizes the significance of understanding the historical evolution of the public in various regions of the world — how it exists (or is limited to exist), what it means, and how it has transformed throughout critical historical junctures. The emergence of present-day data publics is inevitably shaped by this old and new legacy of local public formations.

26 Jonathan Gray, "Three Aspects of Data Words," *Krisis: Journal for Contemporary Philosophy*, no. 1 (2018): 5–17, https://archive.krisis.eu/three-aspects-of-data-worlds/.

PUBLIC

PUBLIC

PUBLIC

BY

AMINA ABDU and ABIGAIL JACOBS

ADMINISTRATION

ADMINISTRATION

ADMINISTRATION

PUBLIC ADMINISTRATION

By Amina Abdu and Abigail Jacobs

Viewpoints from Within the Datafied State

The federal government in the United States comprises elected representatives, like the president and members of Congress, and the federal agencies that do the work of implementing public policy, also known as *public administration*. Unlike members of Congress, agency officials are appointed rather than elected to work in agencies such as the Food and Drug Administration, the Environmental Protection Agency, or the Federal Trade Commission. In the early 20th century, public administration in the US was envisioned as the apolitical implementation of the state's will, as expressed by the legislative branch.[1] However, the work of agency officials has since come to be understood as political work in its own right. Thus, at the core of public administration's history in the United States is an ongoing contest over the legitimacy of the administrative state: while the authority of legislators to make political decisions comes from democratic elections, agency officials must look elsewhere.

Yet to make and administer policy, government agencies rely on the trust they foster. The efficiency and efficacy of agencies thus depends on their perceived legitimacy. The US government's strategies to legitimate agency authority has repeatedly shaped the meaning of public administration and, conversely, changes in public administration have raised new questions of legitimacy. Recently, public administration has once again taken a new shape as agency officials turn to datafication, using data-driven algorithmic

1 Nicholas Henry, "Paradigms of Public Administration," *Public Administration Review 35*, no. 4 (1975): 378–386, https://doi.org/10.2307/974540.

systems to automate parts of the policy implementation process previously performed by agency experts. Agencies tout the promise of more objective and efficient government administration — while algorithmic systems also present risks of bias, errors, and opacity — all with uncertain consequences for legitimacy.

We examine the meaning of increasing datafication in public administration within the longer history of US agencies' efforts to pursue legitimacy. Two strategies in particular have been central to the American administrative state's pursuit of legitimacy: minimizing arbitrariness and expanding accountability.[2] This historical struggle is newly salient in the datafied state, in which the locus of administration is moving from government officials to data-driven technologies raising familiar issues of legitimacy, efficiency, arbitrariness, and accountability. While datafication can be seen as one in a long history of efforts to strengthen the legitimacy of the administrative state, it also runs the risk of undermining this hard-fought legitimacy.

Legitimacy in the American Administrative State

In 1887, the Interstate Commerce Commission (ICC) was established to regulate abuse of monopoly power in the railroad industry. The ICC was the first independent federal agency, and in many ways, the beginning of the administrative state, as the US government subsequently established other permanent agencies to take on the work of public administration. As the administrative state grew in the early 20th century, legal scholars began to debate the legitimacy of these agencies under the US Constitution. As Lisa Bressman describes, early paradigms of public administration were primarily concerned with the threat of arbitrary decisions — decisions based on

2 Lisa Schultz Bressman, "Beyond Accountability: Arbitrariness and Legitimacy in the Administrative State," *New York University Law Review* 78, no. 2 (2003): 461–556, https://www.nyulawreview.org/wp-content/uploads/2018/08/NYULawReview-78-2-Bressman.pdf.

agency officials' individual whims and biases rather than reasoning.[3] The administrative state initially pursued legitimacy through efforts to minimize arbitrariness in the implementation and execution of the "objective" policies dictated by Congress. The federal government sought to mitigate the threat of arbitrariness first through detailed legislative directives and then, when this proved inefficient during the New Deal expansion of the federal government, through appeals to scientific objectivity. Finally, in 1946, Congress enacted the Administrative Procedure Act (APA) to standardize agency procedures. These strategies sought to constrain agency discretion in order to produce consistent, predictable, and fair decisions. By the 1970s, however, idealism about objectivity waned, and the administrative state shifted its focus from minimizing arbitrariness and discretion to expanding agency *accountability.*

Accountability requires not just transparency but a richer relationship that Mark Bovens defines as "a relationship between an actor and a forum, in which the actor has the obligation to explain and justify his or her conduct, the forum can pose questions and pass judgment, and the actor might face consequences."[4] The administrative state sought accountability through multiple channels. In the 1970s and '80s, lower courts and, eventually, the US Supreme Court began to require that federal agencies provide justifications for their decisions. This not only enabled greater public accountability through transparency requirements, but also encouraged consistent decision-making by demanding reasoning around deviations from prior judgments. Agencies pursued legitimacy through expanded public participation, like notice-and-comment rulemaking. Defined in the APA, notice-and-comment requires that agencies notify the public of proposed rules and allow the public to submit comments before issuing a final rule. In theory, notice-and-comment relied on the same method as elected officials: ideals of democratic representation and inclusion. However, public

3 Bressman, *"Beyond Accountability."*

4 Mark Bovens, "Analysing and Assessing Accountability: A Conceptual Framework," *European Law Journal* 13, No. 4 (2007): 447–68, https://doi.org/10.1111/j.1468-0386.2007.00378.x.

participation proved challenging. Reading and responding to comments imposed an efficiency burden on agencies and many questioned whether certain groups were being systematically over- or under-represented in participatory political processes. Gradually, beginning with the Reagan administration, the administrative state has pursued accountability by bringing the administrative apparatus under the control of the president, who is subject to the consequences of public opinion. These three mechanisms — transparency requirements, public participation, and the possibility of consequences — reflect key dimensions of accountability, each of which have been used to support the legitimacy of the American administrative state.

From the Administrative State to the Datafied State

In the datafied state, the administration of policy is increasingly tasked to automated decision-making systems. Such algorithmic technologies are frequently introduced and justified as a way to ensure more efficient, more accurate, and less arbitrary decision-making.[5] However, critical scholars note that these technologies pose a significant threat to accountability, transparency, contestability, and public participation.[6] Now with algorithmic systems as the focus, the fight for the legitimacy of the administrative state is being rehashed, with algorithmic systems being used to pursue familiar goals of efficiency, arbitrariness, and accountability but also raising novel challenges, such as the opacity of algorithmic systems, the reduced role of deliberation and reasoning, and the displacement of agency expertise.

5 For more on the promise of efficiency and objectivity, see Ludmila Costhek Abílio and Carolina Cruz, "Bureaucracy," *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

6 Danielle Keats Citron, "Technological Due Process," *Washington University Law Review* 85 (2007): 1249–1313, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1012360#.

Datafication and Arbitrariness

Discussions around government algorithms have centered on many of the same ideals as those pursued by the administrative state: legitimacy, once again contested on the grounds of arbitrariness, accountability, and efficiency. Proponents of administrative algorithms highlight their potential to advance legitimacy, often pointing to increased efficiency and reduced arbitrariness. Critics, however, contend that algorithmic systems in fact pose unique challenges to non-arbitrariness, while simultaneously undermining accountability. Echoing early administrative goals with appeals to efficiency, Madalina Busuioc characterized how "the promise of efficient, low-cost, or 'neutral' solutions harnessing the potential of big data has led public bodies to adopt algorithmic systems in the provision of public services."[7]

In addition to promises of improved efficiency, proposed benefits of algorithmic technology in government include reduced bias, increased accuracy, and increased consistency.[8] In theory, algorithmic technology has the potential to reduce inconsistencies and the place of individual discretion or whims of agency officials through the implementation of rules.[9] In other words, government algorithms promise to mitigate the arbitrariness of decisions made within the federal agencies that use them. Indeed, central to the promise of algorithmic decision-making is the veneer of objectivity and political neutrality[10] pursued throughout the administrative state's long-standing efforts to establish its own legitimacy. In particular, the turn toward automated systems reflects a push toward what Lorraine Daston and Peter Galison call *mechanical objectivity,* a vision of objectivity that suppresses human judgment in favor of rules.[11] Prizing mechanical objectivity reflects a vision of administration where arbitrariness is seen as driven by human

7 Madalina Busuioc, "Accountable Artificial Intelligence: Holding Algorithms to Account," *Public Administration Review* 81, no. 5 (2021): 825–36, https://doi.org/10.1111/puar.13293; See also Virginia Eubanks, *Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2017); Andrew G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement* (New York: New York University Press, 1017); Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, 2016).

8 David Freeman Engstrom, Daniel E. Ho, Catherine M. Sharkey, Mariano-Florentino Cuellar, "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies," *SSRN Electronic Journal*, 2020, https://doi.org/10.2139/ssrn.3551505.

9 Bernard W. Bell, "Replacing Bureaucrats with Automated Sorcerers?" *Daedalus* 150, no. 3 (July 2021): 89–103, https://doi.org/10.1162/daed_a_01861.

10 Madalina Busuioc, "Accountable Artificial Intelligence: Holding Algorithms to Account," *Public Administration Review* 81, no. 5 (2021): 825–36, https://doi.org/10.1111/puar.13293.

11 Lorraine Datson and Peter Galison, *Objectivity* (Princeton: Princeton University Press, 2010).

problems of bias, error, and inconsistency — rather than a lack of reasoning or moral justification.

However, critics of government algorithms and algorithms used in other high-stakes settings have highlighted that algorithmic decision-making systems both reproduce and exacerbate the human bias they purports to solve[12] as well as simply make mistakes.[13] A. Feder Cooper, Jonathan Frankle, and Christopher De Sa highlight, in particular, that randomness is necessarily built into machine learning models, which presents a legal obstacle to non-arbitrariness.[14] Moreover, objectivity is neither achieved nor achievable: systems used to administer policy necessarily quantify social goals and encode values.[15] Algorithmic systems, like human administrators before them, can never implement policy without themselves making policy decisions. Nor can they fully eliminate values or arbitrariness from these policy decisions. Thus, like previous iterations of the administrative state, their legitimacy must rest on their accountability to the public.

Datafication and Accountability

Datafication reshapes accountability in important ways. The quantification implicit in datafication promises to promote accountability through transparency and public participation alike. Quantification in government, such as cost-benefit analysis, promises to make political decisions more visible through the communication of verifiable numbers.[16] Meanwhile, quantification as an instrument of dispute enables stakeholders and publics to participate in democratic debate and challenge powerful institutions in a common and authoritative language.[17] Perhaps as a result of these features,

12 Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104, no. 3 (June 2016): 671–732; Eubanks, *Automating Inequality*.

13 Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, Andrew Selbst, "The Fallacy of AI Functionality," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency,* FAccT '22 (New York, NY, USA: Association for Computing Machinery, 2022), 959–72, https://doi.org/10.1145/3531146.3533158.

14 A. Feder Cooper, Jonathan Frankle, and Christopher De Sa, "Non-Determinism and the Lawlessness of Machine Learning Code," in *Proceedings of the 2022 Symposium on Computer Science and Law*, CSLAW '22 (New York, NY, USA: Association for Computing Machinery, 2022), 1–8, https://doi.org/10.1145/3511265.3550446.

15 Abigail Z. Jacobs and Hanna Wallach, "Measurement and Fairness," in *Proceedings of the 2021 ACM Conference on Fairness*, Accountability, and Transparency, FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 375–85, https://doi.org/10.1145/3442188.3445901.

16 Cass R. Sunstein, *The Cost-Benefit State: The Future of Regulatory Protection* (Chicago: American Bar Association, 2002).

17 William Deringer, *Calculated Values: Finance, Politics, and the Quantitative Age* (Cambridge, MA: Harvard University Press, 2018).

accountability, inside and outside of government, has increasingly been conceptualized and implemented in quantitative terms, often perceived as transparent, objective, and universal.[18]

Discussions around accountable technologies in administration have focused on transparency as a means to accountability.[19] Yet accountability is solved by neither transparency[20] nor explainability, alone.[21] Contestability mechanisms that allow the public to meaningfully shape and challenge systems must be intentionally designed.[22] Previously, domain experts, policymakers, and publics were involved in developing and executing policies, where technologies are increasingly taking this role. However, that expertise is now overlooked in these discussions. The development of technologies does not eliminate judgment, but instead threatens to displace it to less visible and less accountable locations, where software is developed away from experts and its developers are, unlike traditional government officials, unaccountable to the public.

Despite the promise of quantification, the lack of accountability afforded by algorithmic decision-making systems is often cited as a primary challenge for their adoption in government. A. Feder Cooper et al. identify a number of obstacles to algorithmic accountability surrounding the attribution of moral responsibility and subsequent susceptibility to consequences.[23] The number and variety of people involved in the machine learning pipeline creates a problem of many hands, obscuring both who is responsible for any given decision within a complex system as well as to whom accountability is owed. Even when the division of labor within a system is clear, issues that are seen as inherent in the coding process, make it difficult to attribute a cause to a given problem while computers provide an unsanctionable scapegoat for human and institutional decisions.[24] Compounding these issues,

18 Wendy Espeland and Berit Irene Vannebo, "Accountability, Quantification, and Law," *Annual Review of Law and Social Science* 3 (December 2007), https://papers.ssrn.com/abstract=1057001.

19 Michele Loi and Matthias Spielkamp, "Towards Accountability in the Use of Artificial Intelligence for Public Administrations," in Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, AIES '21 (New York, NY, USA: Association for Computing Machinery, 2021), 757–66, https://doi.org/10.1145/3461702.3462631; Bernard W. Bell, "Replacing Bureaucrats with Automated Sorcerers?" *Daedalus* 150, no. 3 (July 2021): 89–103, https://doi.org/10.1162/daed_a_01861.

20 Joshua Kroll, Joanna Huey, Solon Barocas, Edward Felten, Joel Reidenberg, David Robinson, and Harlan Yu, "Accountable Algorithms," *University of Pennsylvania Law Review* 165, No. 3 (January 2017): 633.

21 Andrew D. Selbst and Solon Barocas, "The Intuitive Appeal of Explainable Machines," *Fordham Law Review* 87 (2019 2018): 1085.

22 Kristen Vaccaro, Karrie Karahalios, Deirdre K Mulligan, Daniel Kluttz, and Tad Hirsch, "Contestability in Algorithmic Systems," in *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing* (CSCW '19: Computer Supported Cooperative Work and Social Computing, Austin TX USA: ACM, 2019), 523–27, https://doi.org/10.1145/3311957.3359435.

23 A. Feder Cooper, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum, "Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22 (New York, NY, USA: Association for Computing Machinery, 2022), 864–76, https://doi.org/10.1145/3531146.3533150.

24 Karen Levy, Kyla E. Chasalow, and Sarah Riley, "Algorithms and Decision-Making in the Public Sector," *Annual Review of Law and Social Science* 17, No. 1 (October 2021): annurev-lawsocsci-041221-023808, https://doi.org/10.1146/annurev-lawsocsci-041221-023808; A. Feder Cooper et al., "Accountability in an Algorithmic Society."

Cooper et al. claim that private technology companies, many of whom are involved in the administrative procurement process, have successfully managed to minimize legal liability further minimizing the possibility of sanctions, one of the key elements of Mark Bovens' definition of accountability.[25] The government procurement process by which agencies acquire algorithmic systems poses a particular challenge to government accountability because agency experts abdicate responsibility for policy decisions to third-party developers who are not subject to the same public participation, deliberation, or oversight.[26] Elsewhere, the algorithmic accountability literature has addressed the lack of transparency and explainability of automated decision-making systems,[27] posing an obstacle to Bovens's requirement of explanation and justification.

Provocations: Unraveling the Datafied Administrative State

With accountability in question, the legitimacy of the datafied state is threatened. To contend with these challenges to accountability and legitimacy, public administration in the datafied state requires a new reckoning with the roles of expertise and public participation. When tools are outsourced or simply developed away from the on-the-ground domains in which they are used, what decisions are being made, where, and by whom, are obscured. Ryan Calo and Danielle Citron note that by giving up the expertise that justifies the administrative state, agencies lose an important claim to legitimacy; experts, policymakers, and the public are excluded from input and contestation.[28]

25 Mark Bovens, "Analysing and Assessing Accountability: A Conceptual Framework1," *European Law Journal* 13, no. 4 (2007): 447–68, https://doi.org/10.1111/j.1468-0386.2007.00378.x.

26 Deirdre K. Mulligan and Kenneth A. Bamberger, "Procurement as Policy: Administrative Process for Machine Learning," *European Law Journal* 13, No. 4 (2019): 773-, https://doi.org/10.15779/Z38RN30793.

27 Citron, "Technological Due Process."

28 Ryan Calo and Danielle Keats Citron, "The Automated Administrative State: A Crisis of Legitimacy," *Emory Law Journal* 70, (2020): 797.

Yet the subtle ways that datafication interacts with the administrative state reveals opportunities for intervention and oversight. When technologies take on the role of public administration, the long-standing tensions underlying (legitimate) public administration are made more interconnected, more obscured, and more interdisciplinary. This complexity points to the need for new ways of understanding accountability and legitimacy in the datafied state.

We point to several examples where we can get theoretical leverage — gaining insight into these interconnected processes, while shedding light on obscured processes and drawing on different domains to see how different stakeholders and technologies interact. We highlight several threads that explore the changing role of expertise and the tensions embedded within public administration in the datafied state. Through these examples, we identify where technologies in public administration can be critically interrogated. In each setting, technical implementations of policy change the role of public administration.

Public Participation

Public participation is a core tenet of accountability in the administrative state, and this participation is increasingly technology-mediated. Some of this mediation is more straightforward. For instance, the US government installed a federal Chief Information Officer under the eGov Act over 20 years ago, enabling increasing digital participation by facilitating online access to government information and services.[29] But we also observe a handoff from the human readers of public comments to automatic sorting and summarizing of public comments.[30] While such text analysis can sound neutral, this sorting — deciding what comments are worth pursuing or prioritizing — is political decision-making, displaced.

29 Laura Stanton, "20 Years of E-Government," *Great Government Through Technology*, October 12, 2023, https://gsablogs.gsa.gov/technology/2022/10/12/20-years-of-e-government/.

30 See Ted Kaouk and Christopher Alvares, "Implementing Federal-wide Comment Analysis Tools," *Federal CDO Council*, https://resources.data.gov/assets/documents/CDOC_Recommendations_Report_Comment_Analysis_FINAL.pdf#page=1

Datafied public participation is not only enacted by the state, as in the federal-wide comment analysis system, but also acts as a key interface to outside stakeholders, especially the most powerful and well-resourced. Consider that even before the age of ChatGPT, public participation was already being disrupted by automatically generated comments: In 2017, the Federal Communications Commission's public comment submission was flooded by millions of fake, automatically generated comments to oppose net neutrality, rendering real comments essentially invisible and precluding meaningful public participation.[31] This was the result of a multi-million-dollar campaign by the broadband industry, disguised as grassroots opposition to net neutrality. This example demonstrates a corruption of traditional accountability mechanisms toward monied interests, undermining public trust and preventing notice-and-comment rulemaking from achieving its goals of democratic inclusion and representation. While this undermining is not new — industry groups vastly dominate public interest groups in these settings[32] — we can also observe that datafication is not only enacted by the state but also onto it. In the datafied state, administration as well as representative democracy itself are threatened.

Contestation, Transparency, and Expertise

We argue that algorithmic systems do not necessarily eliminate, but rather transform and displace agency expertise — both scientific expertise and lay expertise are replaced by technical expertise.[33] This transformation assumes that this substitution is appropriate — and moreover undermines the administrative state as the government becomes unable to oversee itself.

31 "Fake Comments: How U.S. companies & Partisans Hack Democracy to Undermine Your Voice," *New York Attorney General's Office*, 2017, https://ag.ny.gov/sites/default/files/oag-fakecommentsreport.pdf.

32 Nicholas Bagley, "The Procedure Fetish," *Michigan Law Review*, No. 118.3 (2019): 345, https://doi.org/10.36644/mlr.118.3.procedure.

33 For more on the displacement of government officials in the datafied state, see Jennifer Raso and Victoria Adelmant, "'Bureaucracy," *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

Yet organizational practices around accountability and the adoption of new technologies offer a useful lens into how different agencies negotiate expertise, public input, and technical advancement. For example, the Census Bureau's adoption of differential privacy in 2020 represents a particular case where the adoption of an algorithmic technology undermined the legitimacy of a US agency. While the Census Bureau embraced many of the best practices identified in the algorithmic governance literature — transparency, stakeholder engagement, and external auditing — the census implementation of differential privacy became a site of significant controversy, ultimately undermining the legitimacy of the Bureau, its processes, and its products.[34] This example illustrates the importance of attending to the shifting nature of expertise in the datafied state. Efforts to engage external experts from non-computer science backgrounds were hampered by technical jargon and a lack of trust in the Bureau's internal technical experts. Moreover, these non-technical experts were implicitly located at the end of the design pipeline to identify errors rather than involved further upstream in the design process, narrowing where public participation, expert oversight, and, ultimately, accountability could occur. An organizational approach to accountability in the datafied state reveals the importance of creating opportunities for meaningful transparency, different kinds of expertise, and opportunities for contestation throughout the entire decision-making process.

Disinvestment Through Datafication

To understand what is at stake, it is essential to understand both sides of datafied administration. Who is administering, and to whom? Who is datafying and who is datafied? A government agency's decision to implement new

34 For more on the 2020 census, see Maria Filippelli, "Public Interest," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

technologies is not merely a mundane decision about process, but also has important implications for the governed.

Virginia Eubanks has powerfully described how the datafication of public administration acts as a means of control.[35] The costs of datafication are not evenly distributed; minoritized groups become more surveilled and more burdened. Especially through state welfare, food, and medical services, datafied systems expose already-minoritized groups to more arbitrariness. Among those already in need of food assistance via the Supplemental Nutrition Assistance Program (SNAP), Eubanks notes that automation widened the gap between white and Black SNAP recipients, revealing the ways that datafication can further marginalize certain segments of the public. The datafied state renders public administration dependent on the datafication of those being administered. In other words, submitting to tracking, surveillance, and datafication becomes a prerequisite to accessing government benefits and services.

Conclusion

In the US, public administration has always been a contested site for legitimacy. The federal government has tried to solve this legitimacy crisis through rules, procedural requirements, public participation, quantification, and, most recently, the automation of policy through computational technologies. Seeing the datafied state as part of this longer history of public administration reveals new challenges for policymakers and the public, but also reveals that some seemingly new challenges are in fact long-standing tensions within the administrative state. Datafication is not a solution to the contested legitimacy of the administrative state, but complicates it. Our

35 Eubanks, *Automating Inequality*.

examples reveal how the effects of datafication are subtle but wide-reaching. Confronting these complications requires identifying meaningful points of legal, technical, social, and political intervention.

We argue that attending to where and when datafication happens can offer insights. For instance, substituting parts of a system — that is, where the administering of policy is replaced by technical tools[36] — may reveal ethical and political challenges. We must examine if, when, and how the datafication of public administration is itself a displacement of important legitimacy mechanisms.[37] Understanding when datafication supports or undermines administrative legitimacy is key, as new technologies also stand to weaken the agencies they intend to help.

Administrative law offers some paths forward. Legal scholars have pointed to the need for technologies to legitimize, not undermine, the administrative state[38] — yet executing this is not trivial. Practically, technologies acquired by governments must go through procurement. Procurement, where technologies replace policymakers and their activities, then becomes an act of policymaking within the datafied state.[39] Thus, regulation of the procurement process itself offers a lever through which technological systems can be regulated. Regulation through procurement offers opportunities to validate system efficacy, reveal intended and unintended system functions, and enable contestation of existing systems.

Fundamentally, understanding datafication — as scholars, citizens, procurers, or policymakers — requires understanding what the technologies are supposed to be doing in the first place. When technologists make design choices in administrative systems, such as enrollment criteria or eligibility for government assistance, they are making governance decisions.[40] Therefore, if we study how social phenomena are encoded in algorithmic

36 Deirdre K. Mulligan and Helen Nissenbaum, "The Concept of Handoff as a Model for Ethical Analysis and Design," in *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber, Frank Pasquale, and Sunit Das (New York: Oxford University Press, 2020), 231–51, https://doi.org/10.1093/oxfordhb/9780190067397.013.15.

37 Jake Goldenfein, Deirdre K. Mulligan, Helen Nissenbaum, and Wendy Ju, "Through the Handoff Lens: Competing Visions of Autonomous Futures," *Berkeley Technology Law Journal* 35 (2020): 835.

38 Danielle Keats Citron, "Open Code Governance," *The University of Chicago Legal Forum* 2008, no. 1 (n.d.): 355–387, https://chicagounbound.uchicago.edu/uclf/vol2008/iss1/9; Calo and Citron, "The Automated Administrative State."

39 Mulligan and Bamberger, "Procurement as Policy."

40 Abigail Z. Jacobs and Deirdre K. Mulligan, "The Hidden Governance in AI," *The Regulatory Review*, July 7, 2022, https://www.theregreview.org/2022/07/07/jacobs-mulligan-the-hidden-governance-in-ai/; Abigail Z. Jacobs, "Measurement as Governance in and for Responsible AI" arXiv, September 12, 2021, http://arxiv.org/abs/2109.05658.

tools, we can reveal how these tools redistribute decision-making power and where they create the underlying potential for harm. Luckily, social scientists can help surface the hidden policy consequences of technical design choices.[41] We need to use these to identify where there are mismatches between stated values, official policies, and technical implementations, not only because these gaps are where harms emerge, but also because they are where legitimacy is most at risk.

We draw from Nicolas Bagley's belief that a "positive vision of the administrative state" is one whose "legitimacy is measured not by the stringency of the constraints under which it labors, but by how well it advances our collective goals."[42] The ongoing struggle for legitimacy of the administrative state is ultimately shaped by datafication. But meaningful accountability that brings both the public and experts back into the datafied state can help move us toward these collective goals.

41 Mona Sloane and Emanuel Moss, "AI's Social Sciences Deficit," *Nature Machine Intelligence* 1, No. 8 (August 2019): 330–31, https://doi.org/10.1038/s42256-019-0084-6; Abigail Z. Jacobs and Hanna Wallach, "Measurement and Fairness," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21 (New York, NY, USA: Association for Computing Machinery, 2021), 375–85, https://doi.org/10.1145/3442188.3445901.

42 Bagley, "The Procedure Fetish," 350.

PUBLIC

PUBLIC

PUBLIC

BY

ANNE L. WASHINGTON and JOANNE CHEUNG

INTEREST

INTEREST

INTEREST

PUBLIC INTEREST

By Anne L. Washington and Joanne Cheung

The public interest is defined by priorities, infrastructures, and other elements that promise a thriving, safe society. The state asks people to share transportation networks, legal sanctions, digital systems, or public space with neighbors in good faith. In exchange for these responsibilities is an assurance of reduced collective harms and accrued benefits for all. While the term suggests a singular approach to a shared social good, the public interest can be elusive.

This keyword essay brings attention to power and embedded social hierarchies lurking in the notion of the public interest. A series of illustrations challenges whether the public interest represents a coherent group of people or even a common set of norms. The paradox of the public interest arises when communal projects explicitly or implicitly prioritize some populations over others. In the 19th century, a portion of New York City that would become Central Park was seized through eminent domain from a thriving community of free African American landowners. In a contemporary example, property records in India released as open data emboldened the technical elite to systematically steal land from the disenfranchised. In its quest to govern through quantification, the datafied state may obscure underlying political disagreements over shared resources, dividing populations as much as defining what they have in common.

After reviewing its origins in public interest law, the essay illustrates assumptions often made when invoking a shared social good. Our essay shows

how the rights of marginalized communities can be easily compromised to accommodate majority preferences. By demonstrating this, we challenge an assumption common in the public interest technology community: that a profit-driven private sector can be outsourced to serve the public without compromise. We question whether the datafied state will acknowledge and avoid these inherent conflicts of interest. Our concluding examples represent the possibility of true collective imagination. We show that sometimes the public interest can be served by centering the marginal to improve the common good.

Not in the Public Interest

Many other goals may conflict with the public interest. Ideological interests may limit resources to only those deemed to be deserving. Political interests seek to accumulate power for one population group, such as the United States housing policy, which provided low-cost mortgages for white people only.[1] Property rights empower those who already own material goods, prioritizing those with inheritances. Financial profit often comes into conflict with the public interest. The Michigan government netted over $60 million through an algorithm aptly named after the legendary monarch who coveted gold — MiDAS, the Michigan Integrated Data Automated System. It did so by assuming most unemployment claims were fraudulent and garnishing taxes of those without work. It is important to note that public interests are not always universal interests. For instance, parents are not the only residents burdened with the expense of municipal schools, because the whole population has an investment in the next generation.

1 Mehrsa Baradaran, *The Color of Money: Black Banks and the Racial Wealth Gap* (Cambridge: Belknap Press, 2019).

The public interest stands out among other interests because it epitomizes the opposite of profiting from or imposing control over others. The idea that the public interest can be singular, without conflicting priorities, suggests a community's needs can be met without compromise by the whims of the wealthy, the political will of the powerful, or the overbearing influence of those who control resources. Paradoxically, a claim that wields the power of the public may in reality mask other interests that serve only the few. A punitive datafied state often makes a service available based on an ability to pay. Ideological, political, property, or financial interests may clash despite an intention to serve everyone.

Background: Public Interest Law

The early 20th century era of monopoly gave rise to the popular conception of the public interest. The United States Supreme Court decisions steadily empowered conglomerate businesses that opposed the Sherman Antitrust Act of 1890. Prominent legal thinker Louis Dembitz Brandeis noted this situation in his address to the Harvard Ethical Society. The oft-cited May 1905 speech recognized that many court cases were effectively one-sided because the wealthiest clients made the strongest arguments. Brandeis observed that serving industrialists was the only career choice for the best legal talent of his generation, leaving no one to advocate for the poor and less powerful.[2] Brandeis envisioned courts where opposing attorneys argued with equal intellectual vigor.

Law in the public interest laid dormant for decades until it was institutionalized through legal education, innovative organizations, and court administrative practices. Legal clinics within law schools were a core driver in

2 Louis Brandeis, "The Opportunity in the Law, Address to the Harvard Ethical Society," *Commonwealth Law Review* 3, no. 1 (1905): 22–30, http://www.minnesotalegalhistoryproject.org/assets/Brandeis%20-%20%20(1905).pdf.

establishing public interest law. Aside from providing free legal advice to local communities, clinics prepare attorneys to work on pro bono cases within firms. The nonprofit sector led the development of innovative legal organizations that would advocate for the rights of the indigent and poor. This was followed by a federal grant program for legal aid organizations in the Legal Services Corporation Act of 1974 (P.L. 93-355). Court administrative systems frequently offer a court-appointed attorney to anyone who cannot afford to pay for legal advice.

Now well-established, public interest law centers the needs of those compelled to use the legal system but who lack financial resources. The public interest goes beyond the intention to transform institutions. Institutions normalize the concept by making services available across many locations and people.

The problem of representing the public interest within the datafied state parallels the concerns of early legal professionals. Technology built in the public interest within a datafied state would need similar institutional support to thrive.

Public Interest Technology for the Datafied State

Technology in the public interest is an attempt to parallel public interest law's success in advocating for the unheard. Public interest technology is also growing at a time of concerns about antitrust and industrial wealth. The history of state technologies, as chronicled by scholars such as James C. Scott and Hazel V. Carby, shows how government quantification structured society from a Roman imperial census to 19th century colonial accounting tables.[3] Scholars of population statistics, like Evelyn Ruppert and Ian Hacking,

3 James C. Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998); Hazel V. Carby, *Imperial Intimacies: A Tale of Two Islands* (New York: Verso, 2019).

monitor the continuing trend of making people visible through classification schemes.[4] The datafied state builds on this digitized wealth to translate social concerns into computational policy.

Modern state governance depends on even more commensuration and enumeration to manage services. State technologies now include automated decision systems, machine learning, and other algorithmic tools mostly purchased from commercial businesses. The datafied state is equipped with tools that can gather and analyze data. Thus, it risks amplifying existing disparities by making sweeping judgments that are designed for efficiency but easily harm thousands of people at scale.

The datafied state builds upon ground truth datasets that obscure decades of political and ideological interests. Consider how cities have been historically zoned, which organizes populations within cities. At its inception in 1916, New York City zoning laws were justified by sanitation concerns, however, in practice, they became a state-sanctioned system for imposing residential segregation. Laws limited immigrants from China into separate and unequal racialized infrastructures now called Chinatowns.[5] African Americans were forced to rent in squalid redlined areas. Early zoning policies are still visible in 2017 statistics showing the life expectancy of residents in East Harlem is 71.2 years, while those mere blocks away, on the Upper East Side, live to 89.9 years.[6] Redlining laws that controlled property investment terms in the 1930s still echo in the Center for Disease Control's Social Vulnerability Index for the COVID-19 pandemic.[7] Any reliance on historic sources must carefully consider how to interpret policy legacies in the data.

4 Evelyn S. Ruppert, *The Moral Economy of Cities: Shaping Good Citizens* (Toronto: University of Toronto Press, 2006); Ian Hacking, "Kinds of People: Moving Target," in *Proceedings of the British Academy,* vol. 151, *2006 Lectures*, ed. by P. J. Marshall (UK: Oxford University Press, 2006), 285–318.

5 Braden Goyette, "How Racism Created America's Chinatowns," *HuffPost*, November 11, 2014, https://www.huffpost.com/entry/american-chinatowns-history_n_6090692.

6 "City Health Dashboard," NYU Langone, 2023, https://www.cityhealthdashboard.com/about.

7 "Not Even Past: Social Vulnerability and the Legacy of Redlining," National Community Reinvestment Coalition and Digital Scholarship Lab, 2020, https://dsl.richmond.edu/socialvulnerability/.

Exclusion for a Majority Interest

Central Park has long been championed as a symbol of public interest. As one of the first public parks in the United States, it represented a removal of 775 acres of land from the private real estate market, provided fresh air and nature, and offered public space that represented, as its architects believed, "a democratic development of the highest significance."

Frederick Law Olmsted, one of Central Park's designers, believed that public parks created a shared culture that overcame economic differences. At its inception, Central Park embodied two conflicting desires: to emulate the elites and to serve everyone else. Central Park materialized the vision of an elite group acting on behalf of the public rather than the public's self-determination of its own needs and desires.

Through Central Park, the elites of New York channeled their aspiration to establish the city's cultural parity with European capitals. As such, New York's new public park would not be modeled after vernacular landscapes embedded within neighborhoods (where "the public" lived), but rather after the formal parks of Europe that imitated the manicured landscapes of the nobility. Central Park was constructed between 1857 and 1876, during the Reconstruction era after the American Civil War. The planners of Central Park seized a portion of its 775 acres from Seneca Village, established after the Civil War as a haven for free Black people who could afford property.[8] Though Central Park was meant for the public, it was created by a small group with special interests (the gentlemen elite) acting on behalf of majority interests (white New Yorkers) at the expense of minority interests (Black New Yorkers and property owners of Seneca Village). At the time, the State of New York required "persons of color" to own at least $250 worth of

8 "Before Central Park: The Story of Seneca Village," Central Park Conservancy, last modified January 18, 2018, https://www.centralparknyc.org/articles/seneca-village.

property to vote. The seizure of Seneca Village to create Central Park meant that African American owners of Manhattan property lost the right to vote. Seneca Village was home to 10 of the 100 eligible Black voters in New York City. The demolition of Seneca Village in the name of "public interest" extinguished a bright spot of representative democracy in the post-Civil War era.

Examples of how elites who promote some vision of a unified public interest can harm the marginalized are not confined to distant history or the United States. Globally, open data advocates argue that if governments create digital material with taxpayer money, everyone should have access to it. But who is actually served? Making public records available through open government data can yield problems when power asymmetries arise.[9] The government of Bengaluru, India, was widely praised for releasing land ownership and title information on the internet in 2006. Bhuvaneswari Raman documented how the newly digitized property records made it easier to surreptitiously steal land owned by those with little technical expertise.[10] The government of Bengaluru gave only the affluent the benefit of the doubt in the face of incomplete or inconsistent documentation. Embedded within public interest are questions about who is excluded and who is not served.

Administrative acts of goodwill by entitled populations who are equipped with state information can quickly escalate to acts of dispossession. Any discussion of public interest is inseparable from a discussion of democracy, which involves the potential conflict between various types of interests: special interests, majority interests, and minority interests.

9 Anne L. Washington, "Uncertain Risk: Assessing Open Data Signals," *Transforming Government: People, Process and Policy* 14, no. 4 (January 2020): 623–37, https://doi.org/10.1108/TG-09-2019-0086.

10 Bhuvaneswari Raman, "The Rhetoric of Transparency and Its Reality: Transparent Territories, Opaque Power and Empowerment," *The Journal of Community Informatics* 8, no. 2 (April 2012), https://doi.org/10.15353/joci.v8i2.3037.

Power for Financial Interests

One hundred years after the creation of Central Park, a small park opened beside the One Liberty Plaza office tower in New York City's Financial District. In 1972, Zuccotti Park was created through a regulation known as Incentive Zoning, which encouraged the creation of public spaces in areas with high real estate value. By permitting developers an additional ten square feet of built space in exchange for one square foot of open space, a public park was created but not owned by the public. The legal constructs that governed Zuccotti Park became prominent during the Occupy Wall Street protests that took place there. After the 2008 financial crisis, outcry over systemic inequality culminated in a public demonstration to symbolize the people's response to financial malfeasance. During the protests, the property owners amended Zuccotti Park's code of conduct to ban "tents, sleeping bags, and lying down," which was then used as grounds to evict protesters. Zuccotti Park remains a green space accessible without physical barriers. However, structural barriers remain, such as the lack of decision-making power restricting the agency of the public to use it as they would like. Similarly, Central Park is open to all, but structural barriers such as the lack of transportation and leisure time prevent everyone from accessing it.

When states enact state authority through commercial businesses, contradictions abound. Technical solutions in the public sector are particularly susceptible to prioritizing financial incentives. However, the dynamic exists in other contexts.

Profit maximization parading as an investment in the public good impacts not only privately owned public spaces like Zuccotti Park. This solution

is the same neoliberal logic that extends to privately owned online public spaces: social media platforms.[11] Social media sites are open and accessible to anyone as long as the public turns itself into users of a private good. In this change of identity, social media injects a profit motive into everyday communication. The public becomes beholden to the policies and whims of private platforms. The short-messaging system previously known as Twitter was core to communication across levels of social hierarchy until it was purchased by a new buyer who changed the platform's affordance. Governments could no longer rely on it to communicate with citizens. The debates of deplatforming, a practice by social media companies of removing users for policy violations, mirrors contentions in physical space, such as when the public interest intent behind Zuccotti Park's public-private partnership[12] and the structural limitations of its private ownership came into conflict during Occupy Wall Street.

Governments have limited means to implement complex digital systems and inherently rely on private businesses, which can confuse financial and efficiency goals. The Australian government paid its contractors a portion of the money recovered through its fraud prevention program, which became known as Robodebt. Although it ended up costing the government millions in legal fees, Robodebt's private contractor received financial benefits for a system that harmed welfare applicants, stressed government employees, and ruined political careers.[13] Conflicting tensions between administrative goals and financial gain can run the risk of limiting the rights of some to enable the convenience of many. Efficiency, as defined by spending the least amount of money to reach optimal objectives, brings benefits at scale and automates harms.

11 Joanne Cheung, "Real Estate Politik: Democracy and the Financialization of Social Networks," *Journal of Social Computing* 2, no. 4 (December 2021): 323–36, https://doi.org/10.23919/JSC.2021.0030.

12 Matthew Bui and Bianca Wylie, "Counterpublics," *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

13 Anne L. Washington, *Ethical Data Science: Prediction in the Public Interest* (New York: Oxford University Press, 2023).

When governments cede their functions to the private sector, the public must bow to the interests of businesses without accountability mechanisms. Aside from profit motives, these non-state actors may want to advance their own political interests, attack certain population groups, or amass personal control over shared resources. From social media platforms to Occupy Wall Street, to Robodebt, the state becomes beholden to the ideology of resource-wealthy individuals who control design and distribution. If it is to survive as a positive tool, public interest must flip from a plutocratic model to a democratic one that prioritizes everyday people over the wealthy. Just as Brandeis envisioned legal advocates for all, a new public interest model centers the margins.

Marginalized Rights in the Public Interest

The promise of the public interest still holds by realizing maximal accessibility rather than averaging to a homogenous, often privileged majority. We are motivated by a theory of change that recognizes the power of the edge cases and the marginal. This section illustrates how to embrace a broader context by imagining truly universal benefits.

Curb Cuts

An inclusive definition of the public is a crucial step. When those who have been systematically excluded begin to advocate for inclusion, the result can create net positive benefits. Angela Glover Blackwell described this dynamic as the "curb-cut effect."[14]

14 Angela Glover Blackwell, "The Curb-Cut Effect," *Stanford Social Innovation Review* 15, no. 1 (2016): 28–33, https://doi.org/10.48558/YVMS-CC96.

Paved sidewalks, though they are public property, were not designed with wheelchair users in mind. The curb prevented wheelchairs from traversing between the sidewalk and street intersections, and as a result, this public property excluded the disabled community. In the 1970s, disability advocates installed a concrete ramp at an intersection in Berkeley without a city permit. This intervention shed light on a persistent structural barrier. In the 1990s, Americans with Disabilities introduced federal legislation that would mandate the curb cut, a ramp cut into the curb of the sidewalk. When initial designs thwarted the needs of the visually impaired, curb cuts eventually included prominent surface bumps to indicate the boundary.[15] Beyond the initial advocates, people rolling luggage, pushing strollers, and even roller skaters all benefit from its built-in affordance. When a normative design excludes particular experiences and needs, changing the norm in service of particular needs can ultimately improve the general affordance of that design.

Community Science

Data collection efforts often reflect the organization or institution more than the people described. The datafied state could emulate how communities collect data salient to their own environment and well-being.

During British Petroleum's Deepwater Horizon oil spill in 2010, information about the damages to the Gulf of Mexico was not accessible to the public. In response to this private sector information blackout, the Public Laboratory for Open Technology and Science organized a data-centered hyperlocal response. They launched open-source balloon mapping kits for concerned communities to document spill-affected sites.[16] This community-driven effort collected over 100,000 images, produced high-resolution maps used by major outlets such as *The New York Times*, and exposed the extent of the oil spill's

15 Bess Williamson, *Accessible American: A History of Disability and Design* (New York: NYU Press, 2019).

16 Jessica Breen, Shannon Dosemagen, Jeffrey Warren, and Mathew Lippincott, "Mapping Grassroots: Geodata and the Structure of Community-Led Open Environmental Science," *ACME: An International Journal for Critical Geographies* 14, no. 3 (September 26, 2015): 849–73.

environmental impact on the public. This successful combination of community science, open technology, and advocacy catalyzed a data-driven environmental justice movement that serves as a watchdog function but also generates solutions. Aerial hyperlocal data collected over New York City's polluted Gowanus Canal uncovered an active spring buried under asphalt.[17] Their evidence eventually led to the Environmental Protection Agency's cleanup and restoration of the canal. In these examples, it is the community, rather than the state, which became datafied — the state benefited from knowledge generated from local datafication by the public and for the public. The contributions on *missing data*[18] and *counterdata*[19] in this collection offer other examples of data collection by communities, but they also show how these efforts challenge the state and what the state could itself be doing.

Summary

Anyone who summons the public interest to justify a position also seeks to claim high moral ground. It may be appropriate to challenge a public interest claim when the options are in total agreement with only the majority or dominant population. The establishment of public interest law, and parallel efforts to manifest public interest technology, seeks to remedy under-resourced groups. A true test of whether something serves the public interest is to see if it has a broad base of support. The public interest will benefit more than one group. The public interest will simultaneously harm more than one group. Better serving the public interest in a more comprehensive sense means to make the implicit negotiations between varied interests explicit. Projects that are truly in the public interest will never neatly align with only one financial, political, property, or ideological interest.

17 "Uncovering a Historic Stream Flowing into the Gowanus Canal," The Public Lab, March 26, 2013, https://publiclab.org/notes/jeff/3-25-2013/uncovering-historic-stream-flowing-gowanus-canal-0.

18 Alessandra Jungs de Almeida, Lauren Klein, and Catherine D'Ignazio, "Missing Data," in *Keywords of the Datafied State,* eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

19 Seyi Olojo, "Counterdata" and Vanessa Massaro, Darakhshan J. Mir, Terrell Mosley, and Nathan C. Ryan, "Counterdata," *Keywords of the Datafied State,* eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

AUTOMATION
AUTOMATION
AUTOMATION

BY

GEORGIA VAN TOORN, CHRIS O'NEILL,
MAITREYA SHAH, and MARK ANDREJEVIC

AUTOMATION
AUTOMATION
AUTOMATION

AUTOMATION

By Georgia van Toorn, Chris O'Neill, Maitreya Shah, and Mark Andrejevic

Introduction

Automation refers to the use of technology, machines, or software to perform tasks without human intervention. The term originates from the Greek word *autos* meaning "self" and *matos* meaning "willing." To be automated, then, is to have a self-willing or self-acting capacity: to have "the power of motion within."[1] Automation has existed as a concept for millennia, with early technical examples including automated water mills for grinding grain and automated looms in textile production. These technologies were designed to reduce manual labor by simplifying daily tasks and increasing the overall productivity of human effort. It wasn't until the Industrial Revolution and the advent of mass manufacturing, however, that automation became a central fixture of capitalist production. The use of machines and assembly lines enabled companies to produce more goods faster than ever before, and as a result, many industries were transformed by automation.

In the US, automation became a topic of public concern after World War II, when automated technologies and mechanized production lines became hallmarks of the Fordist manufacturing process. In this context, automated technologies became associated with the organization of human labor and the large-scale displacement of routine task-intensive jobs by machines. The cultural meaning of automation evolved to include the belief that machines would replace human workers and lead to increased feelings of alienation and dissatisfaction in their jobs.

1 Cecily Devereux and Marcelle Kosman, "Introduction: Our Automated Bodies/Our Selves." *ESC: English Studies in Canada* 42, no. 1 (2016): 1–20, https://doi.org/10.1353/esc.2016.0003.

During the 1960s and 1970s, automation became a catalyst for change in the government and military sectors. As the growing power of digital computers opened new paths for the automation of *logical* and not just *physical* processes, technology came to be understood as a potential solution to the shortcomings and inefficiencies of state bureaucracy. The dream of automation has been key to the allure of the datafied state — automation promises speed, efficiency, and precision within a state apparatus often framed as slow, cumbersome, and beset by inertia. New infrastructures, political discourses, and public institutions emerged around notions of digital governance,[2] e-government,[3] and the virtual state,[4] reflecting the growing role of automated, data-driven technologies in shaping public policy and decision-making processes. But automation does not merely automate existing processes of state governance. In conjunction with techniques of the datafied state, like predictive analytics and biometric technologies, automation has produced a new and far-reaching re-organization of state power and elicited new forms of contestation and resistance.

Automation in the Contemporary Datafied State

In the last decade, big data and artificial intelligence have further transformed the way governments collect, analyze, and use data to automate decision-making processes in various fields, from health care and social services to transportation and policing. Increasing numbers of people now interact with the government online through digital platforms and mobile apps, such as the Mobile Passport Control app through which travelers entering the United States can upload their travel documentation prior to

2 Patrick Dunleavy, Helen Margetts, Simon Bastow, and Jane Tinkler, *Digital Era Governance: IT Corporations, the State, and e-Government*, 1st edition (OUP Oxford, 2008); Michael E. Milakovich, *Digital Governance: New Technologies for Improving Public Service and Participation* (New York: Routledge, 2011), https://doi.org/10.4324/9780203815991.

3 Robin Gauld and Shaun Goldfinch, *Dangerous Enthusiasms: E-Government, Computer Failure and Information System Development* (Dunedin: Otago University Press, 2006).

4 Jane E. Fountain, *Building the Virtual State: Information Technology and Institutional Change* (New York: Brookings Institution Press, 2001), https://www.brookings.edu/book/building-the-virtual-state/.

arrival. These interactions leave behind digital traces, or so-called data exhaust.[5] These and other forms of data are analyzed to gain insights into citizens' behaviors, preferences, and needs, as well as inform policy making, service delivery, and resource allocation decisions.

Today, automated data collection and processing is an integral part of modern statecraft. For centuries, governments have sought to collect and analyze information about their populations, territories, and resources in order to exert power and control over them.[6] The classification and counting of populations by nation-states beginning in the 19th century can be seen as providing a broad historical context for understanding contemporary forms of biopower, which encompasses both the disciplining of the individual and the regulation of the population.[7] Yet the state's ability to collect information and exercise control underwent a significant transformation with the advent of modern computing. Initially, digital technologies, and more recently, big data analytics have expanded the state's surveillance capabilities, giving it unprecedented power to monitor and influence a vast range of activities, from communication and movement patterns to financial transactions and the use of government services.[8]

Data analysis is now so central to the workings of state bureaucracy that the objectives of government are increasingly shaped by the affordances of big data. No longer limited to the simple assessment of citizens' welfare status, data systems are increasingly used to anticipate behavior.[9] Brian Massumi has described this technique as an "operative logic of preemption."[10] Preemptive systems produce a logical reason to believe something about future behavior, constructed in a way that can justify intervention. The probabilistic nature of these models extends indefinitely into an unknowable and hence manipulable future. A predictive policing model, for

5 "The Power Of Data Exhaust," *TechCrunch*, May 26, 2013, https://techcrunch.com/2013/05/26/the-power-of-data-exhaust/; Melissa Gregg, "Inside the Data Spectacle," *Television & New Media* 16, no. 1 (January 1, 2015): 37–51, https://doi.org/10.1177/1527476414547774.

6 Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, (Cambridge, Massachusetts: Harvard University Press, 1998).

7 Taina Bucher, *If...Then: Algorithmic Power and Politics* (New York: Oxford University Press, 2018).

8 Sarah Valentine, "Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control," *Fordham Urban Law Journal* 46, no. 2 (January 1, 2019): 364.

9 Marc Schuilenburg and Rik Peeters, *The Algorithmic Society: Technology, Power, and Knowledge* (Abingdon and New York: Routledge, 2022).

10 Brian Massumi, "The Primacy of Preemption: The Operative Logic of Threat," in *Ontopower: War, Powers, and The State of Perception* (Durham: Duke University Press, 2015).

example, might identify a "pattern" of crime in a particular "high-risk" neighborhood, which in turn legitimates an increase in the number of police officers stationed in that neighborhood. If these police go on to make an above-average number of arrests, then the high-risk categorization of the model becomes "justified." But such a logic is self-fulfilling. It tends toward ever more state intervention in the lives of citizens, the over-policing of marginalized groups, and the perpetuation of social inequality. So-called predictive policing purports to use machine learning systems to anticipate groups at risk of committing crimes, in the absence of actual evidence of crime.

In recent decades, there has been a significant shift toward the use of big data in governmental sectors that handle the provision of social welfare. Public resources are increasingly distributed via profiling, classification, and risk prediction algorithms that use data to classify citizens, assess their eligibility for social assistance, and monitor behaviors at an individual and population level.[11] This shift toward a data-driven welfare state aims to use data to gain a better understanding of citizens' needs and behaviors. This enables more targeted interventions guided by purportedly neutral bureaucratic tools rather than potentially biased human decision makers. However, automated social profiling, also known as citizen scoring, is not without risks. These methods have resulted in discriminatory outcomes for already marginalized groups due to skewed datasets, technical errors, and, in some cases, administrative misfeasance on a system-wide scale.[12] Technology researchers have established that automated decision systems reproduce the biases and discrimination that exist within society, thereby exacerbating patterns of inequality rather than mitigating them.[13] At a systemic level, predictive analytics are transforming the logic of welfare from one based on socialized risk pooling to an individualized approach, where social welfare is

11 Lina Dencik, Arne Hintz, Joanna Redden, and Harry Warne, "Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services," December 2018; Joanna Redden, Lina Dencik, and Harry Warne, "Datafied Child Welfare Services: Unpacking Politics, Economics and Power," *Policy Studies* 41, no. 5 (September 2, 2020): 507–26, https://doi.org/10.1080/01442872.2020.1724928.

12 Philip Alston, "Digital Technology, Social Protection and Human Rights: Report," October 1, 2019, https://www.ohchr.org/en/calls-for-input/digital-technology-social-protection-and-human-rights-report.; Valerie Braithwaite, "Beyond the Bubble That Is Robodebt: How Governments That Lose Integrity Threaten Democracy," *Australian Journal of Social Issues* 55, no. 3 (2020): 242–59, https://doi.org/10.1002/ajs4.122.

13 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor,* (New York: St. Martin's Press, 2017); Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, UK: Polity Press, 2019); Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press).

distributed based on individual risk profiles and personalized assessments.[14] There are concerns that the rise of automated decision-making may erode the core principles of the welfare state project, such as promoting social equity, reducing poverty, and upholding values of fairness, impartiality, and due process.

Emergence and Interaction

The specifics of how automation is implemented in different countries and periods may vary, but there are certain general tendencies. These tendencies can help us understand the inherent tensions that arise in discussions of automation and its impact on society. A general feature of automated systems is the existence of an internal dynamic of self-(re)production, or what Jessica Riskin calls emergence.[15] The concept of *emergence* suggests that life and consciousness, including artificial forms of intelligence, do not exist solely within a specific substance or machine, but rather arise as properties of a physical system that reaches a certain level of complexity. A system is autonomous to the extent that the regularities it generates arise from conditions internal to the system. Riskin's dialectic moreover provides a useful way to conceptualize a number of tensions and contradictions that characterize the contemporary datafied state. Arguably, the most essential inconsistency is that, despite presenting itself as impartial and objective, the datafied state establishes new types of regulation, monitoring, and control through automation. The benefit of automation is typically framed in terms of the ability to deliver better administrative outcomes by introducing efficiency, optimization, and speed into existing bureaucratic procedures.[16] But automation within the state does not simply replace what were once human actions and decisions with their machinic equivalent. Rather, automation

14 Torben Iversen and Philipp Rehm, *Big Data and the Welfare State: How the Information Revolution Threatens Social Solidarity* (Cambridge: Cambridge University Press, 2022).

15 Jessica Riskin, *Genesis Redux: Essays in the History and Philosophy of Artificial Life* (Chicago: University of Chicago Press, 2007).

16 Karen Levy, Kyla E. Chasalow, and Sarah Riley, "Algorithms and Decision-Making in the Public Sector," *Annual Review of Law and Social Science* 17, no. 1 (October 2021): 309–344, https://doi.org/10.1146/annurev-lawsocsci-041221-023808.; Agneta Ranerup and Helle Zinner Henriksen, "Digital Discretion: Unpacking Human and Technological Agency in Automated Decision Making in Sweden's Social Services," *Social Science Computer Review* 40, no. 2 (April 2022): 445–61, https://doi.org/10.1177/0894439320980434.

begets automation, in what Mark Andrejevic terms a cascading logic of automation[17] — once a particular task or decision is automated, it tends to generate new tasks and decisions that must also be automated, leading to a self-reinforcing cycle of technological expansion.

Antoinette Rouvroy, Thomas Berns, and Liz-Carey Libbrecht, in their work on algorithmic governmentality, argue that this form of rule "circumvents and avoids reflexive human subjects, feeding on infra-individual data which are meaningless on their own, to build supra-individual models of behaviors or profiles without ever involving the individual, and without ever asking them to themselves describe what they are or what they could become."[18] By "infra-individual" data, we understand the disaggregation of subjects into various types of data points, whose relations with one another can generate patterns that do not necessarily do justice to the subject in its entirety. A certain set of attributes, when compared with similar attributes across a population, might augur a credit or security risk, which can then be acted upon in ways that bypass subjective modes of response. A different set of attributes could mark an individual as a suspect of welfare fraud, leading the state to respond without allowing them to present their viewpoint or explain their version of events. In the automated distribution of care, statistical measures of impairment are often employed to assess eligibility for health and disability services. Yet this method overlooks subjective, embodied experiences of disability, and in doing so hinders people's capacity to participate as equal epistemic agents in comprehending their own lived experience.

In other words, when automated systems are unleashed on the world, their interaction effects change how citizens relate to the state. Automated governance not only prevents citizens from participating in decision-making processes, but, in fact, undermines the very notion of a political subject.

17 Mark Andrejevic, *Automated Media* (New York and Abingdon: Routledge, 2020).

18 Antoinette Rouvroy, Thomas Berns, and Liz Carey-Libbrecht, "Algorithmic governmentality and prospects of emancipation," *Reseaux* 177, no. 1 (October 2013): 163–96, https://doi.org/10.3917/res.177.0163.

Historically, radical political theory has envisaged a revolutionary subject able to dismantle existing power structures through "some combination of will, position and knowledge alongside a certain force of history."[19] Yet as Justin Joque argues, "today, the revolutionary subject is beset simultaneously by an algorithmically fragmented reality and an intensely managed digital control."[20] As Rouvroy, Berns, and Libbrecht put it, these forms of control rely upon "a certain type of (a)normative or (a)political rationality founded on the automated collection, aggregation and analysis of big data to model, anticipate and preemptively affect possible behaviours."[21] This analysis of automated platform governance points to (a)subjective modalities of control: automated systems rely on unlimited data collection (any data point is potentially relevant to the extent that it can form patterns with other data points) deployed according to logics of preemptive intervention, acting on individuals before they can act themselves. Risks are detected in advance of their actualization. The route to control is not via subjective agency but through external interventions in real time and modulations in the environment or milieu, at the level of the "rules of the game."[22]

Biometric Data and Automated Governance

The datafied state is increasingly operationalizing the logic of automated governance through biometric technologies — systems that analyze patterns in physical, biological data to identify trends or project behaviors. The implementation of facial recognition technology at international borders, for example, enables governments to regulate the movement of people across borders, using facial information to recognize and flag people perceived as posing a security threat. Governments use biometric identification systems to verify the identity of people receiving public benefits, such as welfare or

19 Justin Joque, *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*, (New York and London: Verso Books, 2022).

20 Joque, *Revolutionary Mathematics*, 16.

21 Rouvroy, Berns, and Libbrecht, "Algorithmic Governmentality," 231.

22 Michel Foucault, The Birth of Biopolitics: Lectures at the Collège de France, 1978-1979 (New York: Picador, 2010), 159–184.

health care, drawing on data including fingerprints and iris scans. Maitreya Shah, one of the coauthors of this essay, has experienced firsthand the uneven impact of biometric technologies. He was denied enrollment in Aadhaar, India's digital ID program, after the biometric technology failed to recognize his iris due to his visual disability. Consequently, he faced significant barriers in accessing financial resources, health care, and government services. As the Indian government started linking Aadhaar with other state programs, many instances surfaced where people with disabilities, similarly situated, were denied crucial benefits.[23] Migrant laborers and rural communities lost access to food grants when Aadhaar's biometric authentication of their fingerprints failed.[24]

Biometric technologies like Aadhaar play a crucial role in defining the limits of citizenship, the "social sorting"[25] of bodies considered worthy of rights and mobility, and discriminating between those who can be seen to belong and those who remain invisible.[26] These technologies redefine the very meaning of citizenship in a world where "human bodies become substantial carriers of information."[27] What is significant here is not just the transformation of human bodies into data, but also the way in which these bodies, from the point of view of the datafied state, stand in for political personhood. The body is tasked with assuming a kind of coherence and order that is no longer available to the datafied, disaggregated political subject. For Avi Marciano, "the direct communication between technologies and bodies, paired with the declining prominence of the mind, renders human communication and negotiation superfluous." This prioritization of body over mind transforms democratic politics: the "employment of biometric technologies … produces mute individuals whose bodies speak for them, and who are not obligated — and sometimes not allowed — to participate, consent, or even speak."[28] Citizens are "stripped of their political status (*bios*) and reduced

23 Smriti Parsheera, "Participation of Persons With Disabilities in India's Aadhaar Project," SSRN Scholarly Paper, (September 2020), https://doi.org/10.2139/ssrn.3700984.; Zubeda Hamid, "Aadhaar a Double Whammy for the Disabled," *The Hindu*, November 21, 2017, https://www.thehindu.com/news/cities/chennai/aadhaar-a-double-whammy-for-the-disabled/article20629931.ece.

24 Human Rights Watch, "India: Identification Project Threatens Rights Ensure Access to Essential Services," *Human Rights Watch*, January 13, 2018, https://www.hrw.org/news/2018/01/13/india-identification-project-threatens-rights.

25 Btihaj Ajana, *Governing through Biometrics: The Biopolitics of Identity* (London: Palgrave Macmillan, 2013).

26 Parsheera, "Participation."

27 Avi Marciano, "Reframing Biometric Surveillance: From a Means of Inspection to a Form of Control," *Ethics and Information Technology* 21, no. 2 (June 2019): 127–36, https://doi.org/10.1007/s10676-018-9493-1.

28 Marciano, "Reframing Biometric Surveillance," 128.

to their biological status (*zoe*),"[29] with deleterious consequences for bodies marked deviant or marginal.[30]

If we acknowledge the idea that automated systems are socially embedded, however, then we ought not exaggerate the power of technology to entirely eviscerate politics. As states pursue the path of automation, they are certain to rub up against the grain of that which resists automated recognition, that is, to encounter political resistance. Globally, there are many initiatives aimed at questioning how automated systems produce knowledge, how they refashion the world in ways that are detrimental to communities and democratic values. Activists are taking action against the datafied, automated state through organizing, direct action, and producing reports and documents that support grassroots resistance.[31] In the UK, the Greater Manchester Coalition of Disabled People is initiating legal proceedings against the Department of Work and Pensions to draw attention to the potential bias in its algorithm for detecting fraud.[32] The Stop LAPD Spying Coalition is working toward "building power toward abolition of the police state," including its arsenal of automated tools that enact forms of police violence.[33] These resistance movements are vital in raising awareness about the potential harm caused by automated systems, and in ensuring that the power of technology is harnessed for the collective good, rather than for state and corporate interests. Activists and civil society groups are advocating for greater accountability in government, and ultimately a more democratic use of technology by challenging the values and injustices built into automated systems and pushing against their limits and inconsistencies.

29 Erin Kruger, Shoshana Magnet, and Joost Van Loon, "Biometric Revisions of the `Body' in Airports and US Welfare Reform," *Body & Society* 14, no. 2 (June 2008): 99–121, https://doi.org/10.1177/1357034X08090700.

30 Georgia van Toorn and Jackie Leach Scully, "Unveiling Algorithmic Power: Exploring the Impact of Automated Systems on disabled People's Engagement with Social Services," *Disability & Society*, 2023, https://www.tandfonline.com/doi/full/10.1080/09687599.2023.2233684.

31 Data for Black Lives (2023) https://d4bl.org/; Fabio Chiusi et al., "Automating Society Report 2020," AlgorithmWatch, October 2020, https://automatingsociety.algorithmwatch.org/; https://d4bl.org/; "Face Off: The Lawless Growth of Facial Recognition in UK policing," Big Brother Watch, May 2018, https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf.

32 "Algorithmic Injustice: An Interview with Rick Burgess from GMCDP," Greater Manchester Law Centre, https://www.gmlaw.org.uk/2022/03/31/algorithmic-injustice-an-interview-with-rick-burgess-from-gmcdp/.

33 Stop LAPD Spying Coalition, "Dismantling Predictive Policing in Los Angeles," May 8, 2018, https://stoplapdspying.org/wp-content/uploads/2018/05/Before-the-Bullet-Hits-the-Body-May-8-2018.pdf.

Conclusion

The datafied state must work to make automation work. This work has often had violent or otherwise harmful effects upon those people and populations who have already borne the brunt of state discrimination and disregard, even as these harms are rearticulated within new logics of governance. To understand the meaning of automation in the datafied state, then, we must do two things. On the one hand, it is necessary to consider how automation serves as an ideal model for the transformation and modernization of the bureaucratic apparatus. On the other hand, we must critically examine how this ideal confronts and is challenged by the messy reality of underfunded services, on-the-ground exigencies, and the irreducible and confounding role of the political.

DIGITAL IDs

DIGITAL IDs
DIGITAL IDs

BY

MARDIYA SIBA YAHAYA and BONNITA NYAMWIRE

DIGITAL IDs
DIGITAL IDs
DIGITAL IDs
DIGITAL IDs

DIGITAL IDS

Mardiya Siba Yahaya and Bonnita Nyamwire

Interconnected and Moderated Bodies

Advocates for digital IDs claim they can provide legal identity to many who lack it, streamline government services, and reduce corruption.[1] Yet digital IDs inherit histories of structural inequalities and reproduce anxieties among the most marginalized.[2] Our contribution contends with these seemingly irreconcilable conditions by weaving together two key arguments. First, biometrics-based digital IDs are data that make people's bodies available for scrutiny at a distance. Second, digital IDs as aggregated datasets serve as a representation of the state, where logics of development and anti-corruption become the justification to collect more data in the pursuit of inclusion.

Digital IDs transform biometric information such as fingerprints, iris scans, and facial features into data. Approaching digital IDs as simultaneously standing in for data *and* the bodies of individuals reveals the conditions of surveillance that disproportionately target marginalized groups.[3] This data-as-bodies approach offers a situated perspective on the implications of the datafied state for the lives of women, gender and sexual minorities, and marginalized ethnic groups. For many among these groups, engaging with the datafied state brings up a range of anxieties — from losing access to systems and services because of failures in registration to the amplification of existing discrimination. Yet, there is also hope — hope to

1 World Bank Group, "Identification for Development: Strategic Framework."

2 Deborah Posel, "Race as Common Sense: Racial Classification in Twentieth-Century South Africa," *African Studies Review* 44, no. 2 (September 2001): 87–114, https://doi.org/10.2307/525576; Paul N. Edwards and Gabrielle Hecht, "History and the Technopolitics of Identity: The Case of Apartheid South Africa," *Journal of Southern African Studies* 36, no. 3 (September 2010): 619–39, https://doi.org/10.1080/03057070.2010.507568; Keith Breckenridge, "Biometric State: The Global Politics of Identification and Surveillance in South Africa, 1850 to the Present," *The International Journal of African Historical Studies* 48, no. 1(2015): 148–150, http://dx.doi.org/10.1017/CBO9781139939546.

3 In this essay, we define marginalized as ethnic, racial, gender, sexual, and religious minorities. Gender and sexual minorities include women, queer, and non-conforming people.

finally achieve legal documentation of one's identity. In between this diversity of lived experiences, marginal citizens must navigate through existing relations of power to confront the limitations of the choices available to them.

Simultaneously, the aggregated datasets of unique digital IDs for all citizens serve as representations of the state; this representation is used in developmental efforts toward efficient and improved service delivery, achieving inclusivity, and tackling corruption.[4] In Zimbabwe, for example, after the Public Service Commission introduced a biometric system and ran an audit in 2020 in collaboration with the World Bank, they found 3,000 so-called ghost workers and removed them from the state's payroll.[5] However, signing up for biometric systems in various countries includes substantial information about and beyond those of the main registrant — the registrant's parents' name, parents' residence, or marriage details and certification,[6] to name a few. This data-as-state approach shows how national identification systems have historically provided states with the power to define acceptable citizen identities — shaping them into machine-readable humans.[7]

We focus on these two approaches to digital ID — data-as-bodies and data-as-state — to demonstrate how groups of peoples' socio-cultural, ethnic, gendered, and religious positionality affects how they engage with the state, how their data is used to inform service provision and delivery, and/or structurally discriminate against them. In our view, current datafied societies are embedded in regimes of monitoring and control, where data is used to make life-altering decisions for people whose data-as-bodies show up well before their actual selves.[8] Thus, throughout our essay, we trace the similarities and differences in historic and current ways identification exists and shows up in the lives of marginalized groups in Africa, the threats of control within datafied states, and a reflection that leaves more questions for further research.

4 Ranjit Singh, "Give Me a Database and I Will Raise the Nation-State," *South Asia: Journal of South Asian Studies* 42, no. 3 (May 2019): 501–18, https://doi.org/10.1080/00856401.2019.1602810.

5 Finbarr Toesland, "African Countries Embracing Biometrics, Digital IDs," *African Renewal*, February 2, 2021, https://www.un.org/africarenewal/magazine/february-2021/african-countries-embracing-biometrics-digital-ids.

6 Quito Tsui and Teresa Perosa, "Digital IDs Rooted in Justice: Lived Experiences and Civil Society Advocacy towards Better Systems," *The Engine Room*, 2022, https://www.theengineroom.org/wp-content/uploads/2022/01/Engine-Room-Digital-ID-2022.pdf.

7 Janaki Srinivasan and Aditya Johri, "Creating Machine Readable Men: Legitimizing the 'Aadhaar' Mega e-Infrastructure Project in India," *In Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers* 1, 101–12. ICTD '13. New York, NY, USA: Association for Computing Machinery, 2013, https://doi.org/10.1145/2516604.2516625.

8 Jasbir K. Puar, "Jasbir Puar: Regimes of Surveillance," interview by Lewis West, *Cosmologics Magazine*, December 4, 2014, audio, https://writology.com/cosmologicsmagazine.

Designing "Machine Readable Humans": The Datafied State's Construction of Identities

Datafied states are the custodians of digital IDs. States have and continue to play the role of creators of "legitimized" identities, the implementers of systematic identification and artifacts that represent people's identities, the interpreters of the data collected, stored, and continuously developed through identification systems. In fact, citizen identities — right from the institutionalization of last names — have been designed as a mechanism to interface with the state.[9] Within and through these interfaces, human identities are converted into data.[10]

For example, during apartheid South Africa, the dompass was instituted in the Pass Laws Act of 1952, which required Black South Africans over 16 years old to carry a passbook at all times.[11] The dompass traced and identified whom a Black person belonged to. "Whom" did not refer to a person's clan, ethnic group, or family but to the white colonizer they worked for. Without the dompass, authorities could not verify whether the Black person had the *"right"* to access *"white"* spaces.[12] A similar requirement was created during the colonial era in Kenya, where indigenes were made to carry passes to access the new capital city of Nairobi.

Today, in our datafied society, whom we belong to in a transnational sense often refers to forms of belonging connected with countries and states based on ethnicity, place of birth, naturalization, or marriage. Yet, within the specific states, "whom" is also used to determine whether you have a claim to citizenship.

9 James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998).

10 Silvia Masiero, "Digital Identity as Platform-Mediated Surveillance," *Big Data & Society* 10, no. 1 (January 2023): 1–5, https://doi.org/10.1177/20539517221135176.

11 "Dom Pass," Digital Innovation South Africa, March 31, 2021, https://disa.ukzn.ac.za/gandhi-luthuli-documentation-centre/dom-pass.

12 Paraphrased from a conversation between Mardiya with a South African acquaintance, while researching histories of IDs.

IDs represent claims to legitimately access certain services, privileges, and liberties. Based on the type of identification document one possesses, the level of available civic services and liberties ranges from high to none.[13] In each case, the state holds the power to interpret the data provided through each identification document or number at their discretion. This also means that when someone does not have access to a legal identity, their freedom of movement and access to basic services is blocked. This is the reality of Nubians and double-registered people in Kenya,[14] as well as refugees in Ethiopia who have faced technical barriers while registering for digital IDs.[15]

In 2018, the government of Kenya enforced the National Integrated Identity Management Scheme, requiring all citizens to register through a biometric identity system which they claimed to be a single point of truth.[16] In March 2023, the government of Kenya relaunched another system, the Unique Personal Identifier (UPI), to register all newborn babies and deaths in the country. The government used the UPI for school registrations, linked to citizens' "identification card, PIN number, National Health Insurance Fund, and Kenya Revenue Authority, as well as identify the individuals in life and in death." [17]

The goal of the UPI, according to the government, was to provide accurate insight and data on the country's population. Similarly, over the recent months in Tunisia, the government has rolled out a series of digital IDs including biometric travel documentations and the mobile ID or an e-identity (e-houwiya) that enables citizens to access government services.[18] Both governments argue that they need to create legible citizens to govern appropriately. At the same time, the Tunisian government argues that biometric identification and IDs need to be implemented by 2024 to fulfill the international civil aviation organizations' mandate for machine-readable

13 Ranjit Singh and Steven J. Jackson, "Seeing Like an Infrastructure: Low-Resolution Citizens and the Aadhaar Identification Project," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 2021): 315:1–315:26, https://doi.org/10.1145/3476056.

14 UC Berkeley International Human Rights Law Clinic and Haki Na Sheria Initiative, "Digital Identity and the Legal Obligation to Conduct a Human Rights Impact Assessment in Kenya," April 2023, http://citizenshiprightsafrica.org/wp-content/uploads/HSI-UCB-Digital-ID-HR-impact-assessments-2023.pdf.

15 Zara Rahman and Sara Baker, "Digital ID in Ethiopian Refugee Camps: A Case Study," *The Engine Room*, 2019, https://www.digitalid.theengineroom.org.

16 Puar, "Regimes of Surveillance."

17 "Government Set to Launch Personal Identifier Portal," *Kenya News*, February 16, 2023, https://www.kenyanews.go.ke/government-set-to-launch-personal-identifier-portal/.

18 Chérif El Kadhi, "Tunisia's Digitization Programs Threaten the Privacy of Millions," *Access Now*, April 27, 2023, https://www.accessnow.org/tunisias-digitization-programs-threaten-the-privacy-of-millions/.

documents. Both the UPI and e-houwiya, like many digital IDs, are required to verify citizens' eligibility for state services, including verifying financial compensation, linking the IDs to other documents such as the national ID and passports. These cases demonstrate that datafied states seek to create interoperable systems to expand legibility *and* machine-readable citizens that can be known from a distance.

Within the datafied state, documentation, numbers, codes, and artifacts created to legitimize a person's belonging to a specific territory and access to services represent machine-readable humans. However, for citizens' bodies to be accessible to the state in ways that make them legible, identities must be crafted along certain parameters determined by governing institutions. The dompass existed to restrict free movement of Black South Africans during apartheid, which meant that the government crafted their identity along the parameters of race.[19] A person's race ultimately determined their interaction with the state and access to public spaces and services. In the datafied state, for humans to be machine-readable, their expressions, complexities, and realities must be limited to specific points, often a patriarchal reconstruction of gendered bodies, colonial demarcation of ethnicity, language, belonging, and ability. The power to be the creator, arbitrator, custodian, and interpreter of people's lives through digital IDs enables the state to have the discretion of what is considered legitimate identity or form of belonging versus illegitimate.

The case of Kenya and Nubians being denied identification has provided a clear case of states' power in determining belonging. The process requires people to provide certain forms of documentation or human verification that may not be available to them in the first place.

19 Posel, "Race as Common Sense."

Recognizable identities created at the state's disposition reinforce violence against minoritized genders. Within most states mentioned throughout this essay, non-conforming gender identities are criminalized. As we will explore further in the next section, the state's role in determining and controlling how people express their complex and fluid identity is enacted through the datafication of the body where a machine-readable human must be quantifiable in patriarchally acceptable ways. Meanwhile, the construction of the datafied body tends to follow the "traditional Western view of personhood as rationality"[20] that encodes people's interconnected, complex, and evolving lives into a set of scientific and mathematical formulas. States echo a logic that claims that their identification systems provide a single point of truth, often at the expense of minoritized groups. Such logic becomes harmful when the lives and identities of people are interconnected, making it possible to disproportionately target entire communities through automated systems that produce generalizations[21] and reconfigure violence and dehumanization.

Monitoring and Control: Cases of Surveillance Within Datafied States in Africa

In March 2022, a Ugandan queer activist based in South Africa stated in a video that she was advised that if she arrives at the Entebbe airport, she will be arrested immediately. While the video was a call to action to picket the Ugandan embassy in Pretoria against legalizing homophobic violence, the activist's story particularly emphasized a version of state surveillance. The Ugandan state is able to take action against the activist because it can identify through her machine-readable documents that she is part of and supports

20 Sabelo Mhlambi, "From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance," *Carr Center Discussion Paper Series*, no. 2020-009 (July 2020), https://carrcenter.hks.harvard.edu/publications/rationality-relationality-ubuntu-ethical-and-human-rights-framework-artificial,2020).

21 Mhlambi, "From Rationality to Relationality."

a criminalized social group. Surveillance and exclusion are part of the architecture of digital identity platforms.[22] The state's ability to provide databases and information that enable profiling and policing with interoperability creates the possibility for surveillance.

Surveillance through interoperability[23] in the operationalization of digital ID systems is evident in the case of Ethiopian refugees who were miscategorized when they were initially registered and were later unable to register for a digital ID due to discrepancies in the system.[24] In Kenya, internally displaced persons who were affected by a severe drought in Northern Kenya that occurred at the same time as the Somali civil war[25] experienced the implications of surveillance when they attempted to register for a national ID, only to find out they had been categorized as refugees, blocking them from accessing national ID cards. Such errors affected multiple communities, and the lack of nuance based on "rationality"[26] restricted access to services for refugees and internally displaced persons.

Surveillance is also organized through social norms, categorizations of acceptable and unacceptable persons or identities and narrative shaping. It "uses such hegemonic norms and narratives to design multiple separations of people into normal/abnormal, good/evil, ally/enemy."[27] In creating these separations, Muslims, ethnic minorities such as Nubians, refugees, and double-registered people, are made[28] foreign by the system. Here Pumla Dineo Gqola points out that identity is performed across boundaries of difference,[29] and people such as educators and social workers who came to represent safe spaces or support the inclusive social development of marginalized groups, become the ones who facilitate monitoring and violence against the communities they are supposed to protect.

22 Masiero, "Digital Identity as Platform-Mediated Surveillance."

23 "Government to Launch Personal Identifier Portal."

24 Zara Rahman and Sara Baker, "Digital ID in Ethiopian Refugee Camps: A Case Study," The Engine Room, 2019, https://www.digitalid.theengineroom.org.

25 "When ID Leaves You Without Identity: The Case of Double Registration in Kenya," *Privacy International*, December 20, 2021, https://privacyinternational.org/video/4412/when-id-leaves-you-without-identity-case-double-registration-kenya; UC Berkeley International Human Rights Law Clinic and Haki Na Sheria Initiative, "Digital Identity and the Legal Obligation to Conduct a Human Rights Impact Assessment in Kenya," April 2023, http://citizenshiprightsafrica.org/wp-content/uploads/HSI-UCB-Digital-ID-HR-impact-assessments-2023.pdf.

26 Mhlambi, "From Rationality to Relationality."

27 Mardiya Siba Yahaya, "What Can Digital Surveillance Teach Us about Online Gender-Based Violence?" GenderIT.org, November 1, 2021, https://genderit.org/feminist-talk/what-can-digital-surveillance-teach-us-about-online-gender-based-violence.

28 Pumla Dineo Gqola, *Female Fear Factory: Gender and Patriarchy Under Racial Capitalism* (Nigeria: Cassava Republic Press, 2022).

Within various African countries, SIM card registrations provide other avenues for increased data collection and surveillance. While SIM cards have long represented a form of digital identity, SIM registration has become an invasive area where the state, through telecommunication companies, promotes the surveillance of its citizens. For example in Uganda, for an individual to register for a SIM card they must present an original national identification card, passport or number, which must be verified by the SIM card-selling officer using a two-step authentication process. In addition, the telecom operator must obtain the photograph of the SIM card applicant. This same situation applies to Nigeria, Ghana, Namibia, and most recently Zambia. Such registration requirements exclude many marginalized groups such as ethnic minorities or migrant workers, and women without ID proof such as birth certificates, needed to obtain a digital ID, which has become mandatory to get a SIM card. Simultaneously, without structural safeguards, the datafied state creates additional databases linking citizens to their mobile number and interpersonal transactions, loosening the boundaries between tracking, identifying, monitoring, screening, and tabulating.

The increased datafication of the people's lives and bodies through SIM cards widens the bounds of how people become legible within a datafied state. The state logic that datafication of people through digitalization will provide single points of truths, including addressing corruption and streamlining service delivery, operates based on conceptions that governance requires people to be made recognizable and legible. However, the implications of surveillance within the datafied state are not evenly distributed. For instance, a gender and sexual minority or migrant worker whose body has been categorized as a threat enforced through public policies and legislature is more likely to be targeted through these interoperable systems that monitor, screen, and analyze their day-to-day interactions. If a person's SIM

29 David W. Tarbet, Michel Foucoult, and Alan Sheridan, "Discipline and Punish: The Birth of the Prison," *Eighteenth Century Studies* 11, no. 4 (January 1978): 509, https://doi.org/10.2307/2737970.

card is linked to their digital ID, biometrics, mobile money transactions, and internet activity, it creates multiple avenues for datafied states to enact harm. Preemptive surveillance to track and shape what a person will do in the future forces ethnic, racial, religious, gender, and sexual minorities to constantly navigate between the choices of enrolling in digital ID systems or opting out to their detriment. Yet by merely engaging in life, social and civic interactions within datafied states and societies, people's data and information is collated, tabulated, tracked and screened, regardless.

Reflecting on the Complicated Fluidity of Individual Autonomy Versus Communal Data

Individual identity, as argued throughout this essay, is moderated by the state. Simultaneously, states make decisions across differences and similarities in how certain groups perform their identities. Data gathered on one person may produce insights on people whose attributes fall into similar categories. For example, when a person is registering for a digital ID, they have to provide verifiable information on their parents and other family members. Similarly, the case of the Ugandan activist provides insights into how decisions are rarely individualized, and made based on social group identities. At the same time, automated immigration decision-making is often based on what people of certain races, locations, and identity "might do." This includes predefined problematic assumptions such as that Muslims are more likely to be engaged in violence extremism or that someone from the Global South is more of an immigration risk than someone from the Global North. All these decisions are encoded in policy, structural designs, and cultural hegemonies reproduced through identification systems and technologies.

While the previous sections have illustrated the differences between the data-as-bodies approach and the data-as-state approach in understanding the politics of digital IDs, we conclude with a reflection of a deeper similarity between them. Both these approaches are fundamentally grounded in a relational view to datafication.[30] This relational view opens up questions such as — what does one's data reveal about a community of people? How does that information create tensions between the individual and the community? How does one tackle a situation where an individual may have consented to data collection, but their data implicates other people "like them" who had no part in that process? What does this mean for a person's individual identity and their right to privacy?

Many African philosophies and practices have evidenced that human engagement, personhood and lives are representations of communities, and eventually flourish through such forms of solidarities. While we do not have specific answers to these questions, they enable us to draw the tensions between autonomy and belonging in the context of datafication. The opposite can also be true, where such relationality provides an opportunity to manifest autonomy through belonging in a participatory approach to designing inclusive data systems. We critique individuation through datafication as a continuation of a history of harms against marginal communities and perpetuation of ongoing forms of violence against them. The simplest way to challenge this process is to ask ourselves whether single points of truths can ever be an accurate approach to govern our complex and fluid communal lives.

30 Salomé Viljoen, "A Relational Theory of Data Governance," *Yale Law Journal* 131, no. 2 (November 2021): 370–781, https://doi.org/10.2139/ssrn.3727562.

PUBLIC
PUBLIC

INTEREST
INTEREST

TECHNOLOGY
TECHNOLOGY

BY
MARIA FILIPPELLI

PUBLIC INTEREST TECHNOLOGY

By Maria Filippelli

There's a recurring conversation I have at every organization where I've ever worked. It begins with a colleague approaching me with a technical issue. It could be that a website is down or a product is glitchy — basically, something is not functioning as expected.

So I troubleshoot with them. I ask what was supposed to happen, like a website being available 24/7. Then I ask if it's happened before. If so, the next question is about the resolution, with the hope that we could repeat that solution. If not, the conversation shifts to me asking about the kind of maintenance in the vendor contract and the contact person listed. The response is often, "I don't know." When I ask to see the contract, service and maintenance terms are minimal if they're listed at all.

Sometimes I'm approached about something a little more personal, like an email regarding a data breach from one of a person's accounts (pick a retail store, hotel chain, or any other place you've had a digital interaction). Unfortunately, it's near impossible to do anything after the fact. The next best thing is to use it as an allegory for better personal and organizational cybersecurity. Especially in the past few years when personal and work devices have become so entwined.

In either situation, the conversation is happening after the fact. As a Public Interest Technology (PIT) practitioner, I often balance fixing problems after something breaks while designing products, policies, and practices that minimize the chances of them happening in the future. PIT practitioners often advise on technologies, reconcile new technologies with legacy technologies, and take the larger technical infrastructure into account while building solutions.

The "public interest" piece of PIT can be an elusive term to define, especially when it comes to technology. At its core, PIT reduces barriers to entry for public services — both for folks accessing that service as well as those operating the service. The datafied state[1] proposes helpful questions about facets of public interest, like procurement and accountability, which have long been parts of my work. However, it's important to note that technologies can also create barriers to entry, like requiring reliable internet access, so sometimes the right tech solution is not tech at all. Whatever the project, PIT involves understanding the problem you're trying to solve and extensive outreach, collaboration, and consensus building.

Throughout this paper, I illustrate some common scenarios at the intersection of the datafied state and PIT, like working with an outdated procurement process and interpreting highly technical concepts for a general audience while balancing various tensions such as timelines, stakeholder expectations, and policies. From a multimodal trip planner to the 2020 Census, I have worked with all levels of government and on every part of a technical project from design to evaluation. The issues are sometimes with the technology, sometimes with the process (outdated procurement language), and sometimes with perception (Are past failures indicative of future ones?). I include some tried and true lessons (asking questions is a

1 "The Datafied State," https://points.datasociety.net/the-datafied-state-a2a7101ba573.

good thing), as well as some topics that are still evolving and will continue to unfold in the near and far future (tensions between highly accurate data and a right to privacy).

An "Open Scope of Work" is Not in the Public Interest

The two guys across the table were not ready to present, fumbling with cords and making jokes the entire time. My fellow interviewers laughed along with faces full of anticipation. I did not feel the same. I couldn't understand why anyone would have such a cavalier attitude on a multi-million dollar contract. But how could they know how much work I put into getting to this point? Hundreds of hours with experts to update the procurement process, with my team drafting requirements, and with partner agencies envisioning "what could be" led to this day.

What got us sitting around the table started several years prior to that day. I worked for a metropolitan planning organization that focused on transportation planning and execution for the region. The organization had long managed commuter programs to reduce peak time roadway congestion. We managed a regional vanpool program with hundreds of vans, an employer-based carpool program, bike lockers, and a few other programs to support commuters. For nearly 20 years these programs were managed by paper and, eventually, a combination of paper and local databases. We were faced with a very common problem in the public sector — how do we digitize a paper-based program making it (in theory) easier for commuters (the public) to access and program leads to manage?

The advent of smartphones and big data had us wondering what was possible. Could we build an application that included all of our programs in one place? This was also before any of the big mapping services had multiple transportation modes available; they were in the vehicle traffic time and sometimes public transportation phase. There was an opportunity to build a new and needed solution.

The first step in the process is research and landscape analysis. We wanted to know which metro areas had something similar to what we wanted, who was thinking about making this happen, and what vendors were in this space. This early in the process, our project was not funded and had no open requests for proposals (RFPs), so we could converse with vendors without any conflict of interest.

Though it was not a crowded space, a few tech vendors were in the multimodal commuter space. They, too, saw the potential in the increased use of smartphone applications and maps. The conversations, akin to an informational interview, followed a similar pattern. I explained what we were trying to accomplish and the vendor would let me know what pieces were developed, what parts were under development, and what parts they couldn't (or wouldn't) do. Through the vendor's limitations, I uncovered what tech concealed and why it is so challenging to protect public interest.

Tech companies want to own their source code, the unique documentation that undergirds their tech products. Their ownership makes their tech proprietary, competitive, and profitable. In addition, I discovered that tech companies prefer little customization or bespoke products, if any. They are thinking of scale — "If we do this in one metro area, how can we replicate it for others?" While bespoke products are an option, they are often too expensive for the public sector. The weeds of the customization part of the

conversation is usually when I heard, “If you just give us a general idea of what you want, we can figure out the details,” or “We like an open scope of work,” or “We don’t like too many specifications.” Vendors that fail to design what works best for your organization and those not willing to have a conversation and collaborate are, by default, creating your tech policy: a policy that likely benefits the tech company more than it benefits the public interest it is intended to serve.

Once we conclude the research, a procurement can be drafted. Procurements include heavy documentation, including scopes of work, budgets, and timelines. Government procurements, at this time and place, were designed where the government owned everything procured. However, this type of procurement would not work for tech and especially tech start-ups, as they would not give up ownership of their source code, and owning source code is not a position governments necessarily want or need. To own source code requires staff to manage it; often licensing a tech product is the correct fit. So contract language needs to be flexible, and in my case — updated.

The process of updating our contracts so that we could license tech products and create customized off-the-shelf products with vendors required … another contract. We needed an expert in intellectual property to help us identify needed changes in the contract and procurement language. The process proved successful as it yielded us terms of ownership over data but not code and set up the organization for all tech products moving forward.

With the research and updated contract language completed, I began drafting the specifications. To date, this is the only time in my career I ever drafted a full technical specification document with purpose, needs, requirements, functionality, and behaviors that we wanted from our tech product. It

is a highly collaborative and detailed process, requiring conversations with anyone who will interact with the product. It's one of those things so foundational to creating tech that it should be mandatory for anyone wanting to work in tech or PIT.

After the RFP is made public and bids are received, they are reviewed and scored. Then, there is an interview for finalists — that led to the conversation with the two joking guys representing their particular tech company. They asked about the usual things, such as flexibility in timelines and budgets, and then asked about the technical specifications. Did they really need to adhere to all of them? Yes. While I felt confident not owning source code, providing a sub-par product to the public is unacceptable.

Everyone is a Stakeholder in PIT

Fast forward nearly a decade from that first procurement process and I'm standing in the hallowed halls of Congress for a day of education and advocacy regarding the 2020 Census. A colleague, from a less technical background, also there to provide census education to advocates and congressional staff asked me, "Do you think you know more about the census technology than we do?" It struck me in two ways. One, I had been hired specifically because I work at the intersection of technology, data, and policy (all intrinsically intertwined, especially in the public sector). Two, I was viewed as an outsider, challenged on my tech, policy, or advocacy chops — I wasn't sure. But I knew there were long-standing, unresolved tensions between DC tech advocates and Silicon Valley … and that comment let me know I was somewhere in the middle.

In the years between my first tech procurement and that fall day in 2018, the world became more digital: data brokers had more information about individuals, the harms of social media were known, and lack of accountability for tech firms persisted.

The 2020 Census underwent a number of updates from the 2010 Census. It would be conducted largely online, so the design of the online form and cybersecurity were two of the biggest concerns. It was also administered in the age of mis- and disinformation (the Cambridge Analytica[2] scandal broke in the spring of 2018). I expected external challenges like limited information from the Census Bureau and apathy from social media companies. The internal challenges were unforeseen.

My work was cut out for me. Not only was there new technology to understand and inform folks about, but it was happening in a crowded space. A network of hundreds of organizations, thousands of stakeholders and volunteers, and many others supported the Census Bureau's work and ensured that the Census Bureau could count everyone "once, only once, and in the right place."[3] There were voices just trying to get the word out about the importance of being counted, advocates trying to ensure the census had proper funding, a presidential election, and a global pandemic vying to shut it all down.

The confluence of all this technology and its implications — census data is used for apportionment, redistricting, and federal funding programs (and that's just the beginning) — required that I learn fast and adapt my approach for different stakeholders. Functioning as a bridge is the best way to describe my work. I bridged technical documentation with everyday language, concerns with facts, and the Census Bureau with stakeholders. The comments about who I spoke with and how I approached my work continued, but the importance of the census outweighed all of that noise.

2 Facebook-Cambridge Analytica Data Scandal, Wikipedia, https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

3 "Counting Everyone Once, Only Once and in the Right Place," Census Bureau Blog Post, November 5, 2018, https://points.datasociety.net/the-datafied-state-a2a7101ba573.

Digitizing Government Forms

Prior to the 2020 Census, the last time the federal government had tried to launch a nationwide website was healthcare.gov in 2013 — it was considered an epic fail.[4] The questions I heard about the census was not *if* it would fail, but what we should do when "the inevitable" happened and folks couldn't get online to fill out their form. The fear was that people would try once, and if they couldn't get online, they wouldn't come back, thus leading to an undercount of the population and communities missing out on critical funding, among other things.

The information the Census Bureau provided to debunk the concerns of failure was slim. I was able to find load capacity reports[5] and not much else. All I could tell people was that the system was performing as expected with some flags, and it would be ready on time. The other piece of information I tried to scrape together was what browsers and devices were used for the testing. The response, a fairly standard "the most up-to-date browsers and devices," was actually insufficient. The assumption that everyone in the US has the most updated phone, laptop, or tablet is incorrect. In fact, I found out many lower-income households used devices outside of the tests.

Beyond that were questions of what the form would look like and how the experience would be. For example, "Can you start the form and then come back and finish it later?" I understood and felt the stress — how could I explain something unseen? It turned out the Census Bureau was keeping the UI/UX under wraps as a matter of security. I argued that not showing people what the form looked like was in itself a security risk. If you don't know what to expect you're more susceptible to fakes. Shortly before the form launched, the Census Bureau agreed to a demo of the online form. I took screenshots and quickly made resource guides for multiple audiences.

4 "The Failed Launch of www.HealthCare.gov," November 18, 2016, https://d3.harvard.edu/platform-rctom/submission/the-failed-launch-of-www-healthcare-gov/.

5 Load capacity tests how many people can access a system at one time before it fails.

Back to the load capacity tests. The system under development started failing their capacity tests near the launch date. In perhaps a lesser-known, but rather extraordinary, story about the 2020 Census is that a backup system[6] was built internally. Originally designed as a backup only, it became the main system that collected the majority of census responses. An excellent PIT lesson: redundancy is resiliency.

Reconciling Data, Technology, and Policy

In December 2019, at a National Academies workshop on the 2020 Census Disclosure Avoidance System (DAS)[7], I watched a presenter grip the lectern and adamantly argue that highly accurate data at low levels of geography should be released publicly. The presenter was concerned that the new DAS would negatively affect census data quality. But highly accurate data at low levels of geography makes for easy database reconstruction, leading to the re-identification of individuals. And the Census Bureau knowingly releasing our individual information is prohibited by Title 13, US Code.[8] How that law is interpreted and executed was cause for much debate leading up to 2020, and it remains a hot button issue today.

The Census Bureau does not release data in a digital or data vacuum. The proliferation of digital tools and low cost of computing power available globally requires consideration of the larger data and technology digital space. The Census Bureau understood this challenge and developed a comprehensive mathematical DAS, unlike any previous decade. Differential privacy, the new DAS framework, became the focus rather than discussing what individual privacy means in this current environment and how we should balance different stakeholder needs.

6 "2020 Census: Initial Enumeration Underway but Readiness for Upcoming Operations Is Mixed," GAO, February 12, 2020, https://www.gao.gov/products/gao-20-368r.

7 "2020 Census Data Products," National Academy of Sciences, accessed December 12, 2023, https://www.nationalacademies.org/our-work/2020-census-data-products-a-workshop.

8 "Title 13 - Protection of Confidential Information," US Census Bureau, https://www.census.gov/about/policies/privacy/data_stewardship/title_13_-_protection_of_confidential_information.html.

Quick-changing technologies rarely align well with slower-moving policymaking, especially in cases where we must protect an entire nation's data. However, anytime we release data it enters a global digital space. Layering mathematical and cybersecurity techniques, in addition to developing policies, are tools in the hands of all PIT practitioners that can ensure our right to privacy.

The Digital Ecosystem

The technology used to collect census responses went beyond what the Census Bureau built. While the Census Bureau was working on their internal systems, advocates and other stakeholders were gearing up to help with outreach, engagement, and to fill gaps in the digital divide. Local governments and advocates employed technologies to help create awareness about the importance of participation. They provided tablets, delivered hot spots to areas with poor internet service, and used text messaging services to send informational links and reminders to respond.

In one meeting about the use of these technologies, I confronted the idea of being tech agnostic. The conversation started by probing if additional technologies might be helpful. Then it flowed to what type of technologies should be put into use. Then someone said, "I don't care which technologies we use, I'm tech agnostic." At this point, I raised my hand and pointed out that perhaps we should not be tech agnostic, as that was risky if we're not considering how a tech company is collecting and sharing an individual's data or if we know their cybersecurity setup (the things we were demanding of the Census Bureau).

The response was a very frustrated "We can't afford to be picky about the technology we use right now." And I could not disagree more. We cannot afford *not* to be picky about the technologies we use, basically ever. Tech products collect data, and once our data is handed over to a product, whether through profile setups, cookies, or other social engineering, we have very little recourse when that data is mishandled or sold to data brokers. Sometimes PIT practitioners have to dole out the tough love and ask the hard questions, and we're not always given thanks for it. It is a fine line to walk, can feel awful, but we can't afford not to do it.

The Future is Even More Complex

As I write this paper, there are two big conversations happening at the intersection of data, technology, and policy that make the understanding of the Datafied State and PIT more complex and urgent.

One is that the federal government is making significant changes to the demographic data it collects across all agencies and departments, not just through the census. Specifically, the US Office of Management and Budget, which coordinates the federal statistical system, is reviewing and developing recommendations to expand the federal race and ethnicity categories currently utilized by the government. In addition, the National Science and Technology Council published the Federal Agenda on LGBTQI+ Equity[9] to assist federal agencies in creating evidence-based policy for the LGBTQI+ communities. Combined, these changes will better reflect the diversity of our nation in our nation's data.

There are requests from census stakeholders to release this data, disaggregated and at the lowest levels of geography possible, to better understand

9 "Federal Agenda on LGBTQI+ Equity," January 2023, https://www.whitehouse.gov/wp-content/uploads/2023/01/Federal-Evidence-Agenda-on-LGBTQI-Equity.pdf

economic and health disparities, among other issues. Simultaneously there is a scramble to understand the impact of generative AI, and regulate AI in general.

These conversations are not happening together.

The increased detail we provide on federal forms, like the 2030 Census, is necessary to ensure equitable distribution of federal funds and enforce civil rights laws, among other issues vital to our health and democracy. While the details of these changes are finalized (expected in summer 2024[10]), the proliferation of generative AI tools continues. AI pulls from a wide range of data sources[11] with little, if any, discrepancy. We need to be careful with detailed data publications, or we run the risk of re-identifying individuals, leaving us susceptible to known and unknown harms.

Remember census DAS? The Census Bureau is still working with stakeholders and conducting a participatory algorithmic design process to balance between quality data and privacy protections. And the procurement process changes? Those can happen anywhere at any time. These processes, policies, and practices reduce and limit harm when implemented in the public interest. Being more intentional with our technologies will limit the harm to ourselves and others.

When I run into former colleagues, they often tell me about something they learned from our time together. It may be that they didn't give their personal data to an app, or they included maintenance in a tech contract — something that lets me know the tensions and hard conversations about PIT and the datafied state are worth it.

10 "Frequently Asked Questions," OMB Interagency Technical Working Group on Race and Ethnicity Standards, https://spd15revision.gov/content/spd15revision/en/faqs.html.

11 "Higher Education and Generative AI: Evolving Lessons from the Field," New America, April 20, 2023, https://www.newamerica.org/pit/events/online-what-chatgpt-tells-us-about-the-future-of-ai/.

OPEN DATA

OPEN DATA
OPEN DATA

BY
MALAVIKA RAGHAVAN*

OPEN DATA
OPEN DATA
OPEN DATA

OPEN DATA

By Malavika Raghavan

Introduction

This essay considers the term "open data" and the evolving role of open data initiatives in modern datafied states. Broadly, open data refers to data that is released in accessible formats for reuse and sharing by anyone. Governments around the world have adopted open data policies to support the timely publishing of data held by public bodies in a reusable, accessible manner using structured or machine-readable formats without restrictions or charges for their use.[1]

The focus of open data initiatives is often centered on form. Supporters of open data champion the release of digital datasets in structured, machine -readable formats to ensure they can be accessed and processed by actors and computers outside the state. In doing so, open data initiatives appear to implicitly acknowledge the role of big data within the datafied state. They also appear to be premised on the existence of citizenry with the ability to analyze and engage with such datasets. Open data emerges as a digital-first response to government accountability and civic engagement.

"Open data" and "open government data" are often used interchangeably in key policy documents and forums. This is despite the fact that the technical and legal arrangements that enable the "opening up" of data can apply to datasets irrespective of whether they originate in the public or private sector.[2] However, the discourse around open data has been overwhelmingly

* The author would like to thank Zainab Bawa for contributions to this piece through conversations about open data initiatives in India. All errors and omissions remain those of the author.

1 Yingying Gao, Marijn Janssen, and Congcong Zhang, "Understanding the Evolution of Open Government Data Research: Towards Open Data Sustainability and Smartness – Yingying Gao, Marijn Janssen, Congcong Zhang, 2023," *International Review of Administrative Sciences* 89, No. 1 (April 28, 2021), https://journals.sagepub.com/doi/10.1177/00208523211009955; Evangelos Kalampokis, Efthimios Tambouris, and Konstantinos Tarabanis, "Open Government Data: A Stage Model," in *Proceedings of the 18th Annual International Conference on Digital Government Research* (dg.o '17: 18th Annual International Conference on Digital Government Research, Staten Island, NY USA: ACM, 2017), 235-246, https://doi.org/10.1007/978-3-642-22878-0_20; Public Resource, "Open Government Data Principles," December 8, 2007, https://public.resource.org/8_principles.html.

2 Maximilian Heimstädt, Fredric Saunderson, and Tom Heath, "Conceptualizing Open Data Ecosystems: A Timeline Analysis of Open Data Development in the UK," in *Proceedings of the International Conference for E-Democracy an Open Government(CeDEM14)*, (2014), 245–256, http://dx.doi.org/10.17169/FUDOCS_document_000000020332; Harlan Yu and David G. Robinson, "The New Ambiguity of 'Open Government,'" *UCLA Law Review* 59, (March 2012), https://doi.org/10.2139/ssrn.2012489.

centered on the role of government and the opening up of public datasets to support societal objectives. This raises an important conceptual distinction that we must confront: that concepts relating to "open government" with "open data" are distinct ideas that may *not* intersect in practice. Harlan Yu and David G. Robinson[3] point out that "open government data" can refer to either of the following:

- *"Open government"* + *"data."* This refers to politically important disclosures that contribute to the openness and transparency of governance, irrespective of whether such data is delivered digitally.
- *"Open"* + *"government"* + *"data."* This data related to the government is made easily accessible, irrespective of its political significance in holding the government to account.

Open data policies adopted by states generally reflect the latter conception: they focus on the form and conditions release of digital datasets to the public. In doing so, they are motivated by broader policy objectives of transparency, participation, or citizen engagement. This raises the question of *how (and whether) we can maintain distinctions between the technologies of open data, and the politics of opening up government data.*

We consider this question by tracing the emergence of ideas of open data, before focusing on India's experience with open data initiatives. India provides a unique vantage point to understand the role of open data in a datafied state, as a country in the Global South that committed early to public sector digitalization and sweeping e-governance reforms.[4] India's approach also reveals the unexpected ways the open data agenda can evolve inside and outside the state. Reflecting on these shifts, the essay concludes by considering future directions and questions for open data.

3 Yu and Robinson, "The New Ambiguity," 181.

4 Radha Chauhan, "National E-Governance Plan In India," United Nations University, May 2009, http://i.unu.edu/media/unu.edu/publication/1377/report414.pdf.

The Emergence of Open Data

Ideas of open data have diverse roots. Researchers have traced ideas of open data to national movements calling for greater transparency in government records management and freedom of information legislation in the mid-20th century.[5] Others trace their emergence to calls from a committee of the United States' National Research Council for an international system of "full and open exchange" of data to improve the scientific understanding of complex global problems.[6] However, the first clear milestone catalyzing principles for open government data came from US-based civil society and internet activists in the 2000s.

In 2007, thinkers and activists gathered in Sebastopol, California, calling for the opening up of government-held data.[7] The gathering included figures such as Lawrence Lessig, Carl Malamud, Aaron Swartz, and Tim O'Reilly, and others from civil society and the free and open source software and resulted in the articulation of eight Open Government Data (OGD) Principles.[8] They consider government data to be open if "it is made public in a way that complies with the principles below:

1. *Complete:* All public data is made available. Public data is data that is not subject to valid privacy, security, or privilege limitations.
2. *Primary:* Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
3. *Timely:* Data is made available as quickly as necessary to preserve the value of the data.
4. *Accessible:* Data is available to the widest range of users for the widest range of purposes.

5 Stefan G. Verhulst, Andrew J. Zahuranec, and Andrew Young, "What the Drive for Open Science Data can Learn from the Evolving History of Open Government Data," *The Conversation*, March 17, 2021, http://theconversation.com/what-the-drive-for-open-science-data-can-learn-from-the-evolving-history-of-open-government-data-156778.

6 Shaida Badiee, Jamison Crowell, Lorenz Noe, Amelia Pittman, Caleb Rudow, and Eric Swanson, "Open Data for Official Statistics: History, Principles, and Implementation," *Statistical Journal of the IAOS* 37, no. 1 (January 2021): 139–59, https://doi.org/10.3233/SJI-200761; National Research Council, *On the Full and Open Exchange of Scientific Data* (Washington DC: National Academic Press, 1995).

7 Simon Chignard, "A Brief History of Open Data," *Paris Tech Review*, March 29, 2013, https://www.paristechreview.com/2013/03/29/brief-history-open-data/; Joshua Tauberer, "History of Movement," in *Open Government Data: The Book* (2014), https://opengovdata.io/2014/civic-hacking/.

8 "Open Government Data Principles"; "Memorandum: Open Government Working Group," October 22, 2017, https://public.resource.org/open_government_meeting.html.

5. *Machine processable:* Data is reasonably structured to allow automated processing.

6. *Non-discriminatory:* Data is available to anyone, with no requirement of registration.

7. *Non-proprietary:* Data is available in a format over which no entity has exclusive control.

8. *License-free:* Data is not subject to any copyright, patent, trademark, or trade secret regulation. Reasonable privacy, security, and privilege restrictions may be allowed."

These principles present a conception of open data as enabling the release of public data[9] at the highest possible level of granularity in a timely, accessible, and machine-processable manner.[10] The focus is predominantly on the **form of the data released:** its completeness and interoperability with different systems that may process it. However, the **conditions of data releases** — that they are non-discriminatory, nonproprietary, and license-free — reflect political and economic choices relating to the "opening up" of government data. They articulate that the government should open up data *freely* and *for free* to the public.

9 The term "public data" is central to the OGD Principles, but a choice was made specifically not to define the term, and focus only on the conditions of its use and management (Public.Resource.Org, 2007b).

10 "Open Government Data Principles."

These conceptions have evolved in subsequent years as governments around the world have operationalized open data initiatives. Some shifts are evident from a review of how different policies articulate their conception of open data (as presented in Table 1).

Table 1: Definitions and Conceptions of Open Data
Source: Author's representation based on cited sources

Institution / Instrument	Conception of Open Data	Key Concepts
Open Knowledge Foundation[11]	Data that can be freely used, modified, and shared by anyone for any purpose.	▪ Free use / Usability ▪ Sharing ▪ Modification
US Open Data Policy (Part I (Definitions))[12]	Publicly available data structured in a way that enables the data to be fully discoverable and usable by end users.	▪ Free use/ Usability ▪ Discoverability ▪ Accessibility (public availability) ▪ Data format (structured)
EU Open Data Directive[13] (Recital 16)	Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used, and shared by anyone for any purpose.	▪ Free use ▪ Sharing ▪ Data format (open format)
UK National Data Strategy (Glossary)[14]	Data that can be freely used, re-used, and re-distributed by anyone, subject only, at most, to the requirement to attribute and share alike	▪ Free use ▪ Redistribution ▪ Attribution
Government Open Data License - India (Part 1. Preamble)[15]	Structured data available in open format and open license for public access and use.	▪ Accessibility (public access and use) ▪ Data format (structured, open format) ▪ Attribution (open license)

Table 1 reveals some commonalities across the definitions of open data presented. They reflect the common premise of open data efforts as enabling accessibility and use of government data by the general public. They are aligned on *the form* of data released (structured, open formats) to support the accessibility and usability of data.

11 Open Knowledge Foundation, *Open Definition: Defining Open in Open Data, Open Content and Open Knowledge* (2023), http://opendefinition.org/.

12 Office of Management and Budget, "Open Data Policy: M-13-13 — Memorandum for the Heads of Executive Departments and Agencies," Executive Office of the President of the United States, May 9, 2013, https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2013/m-13-13.pdf.

13 Council of the European Council, & European Parliament, "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)," *Official Journal of the European Union* 172, no. 56 (2019).

14 Department of Digital, Culture, Media and Sport, "National Data Strategy," GOV.UK, December 9, 2020, https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy.

15 Ministry of Electronic and Information Technology, "Government Open Data License," 2017.

However, some shifts are seen in terms of *the conditions* safeguards, to which open data are subject. The definition in UK's national data strategy, and the license arrangements under India's National Data Sharing and Accessibility Policy (NDSAP), refer to attribution prior to sharing open data. In the Indian context, the requirement for an open license to access open data (and adherence to its conditions) is justified by the need to ensure "such data is not misused or misinterpreted."[16] Open licenses are also commonplace in most other OGD programs. While this reflects the growing understanding of risks of data sharing, it is nevertheless in contrast to the license-free and non-discriminatory (or registration-free) vision under the OGD Principles.

This broad framing sets the scene for a deeper consideration of the extent to which technological operationalization of open data intersects with political and social considerations within the Indian context.

Open Data in India

Formal open data efforts in India have been driven by government actors, emerging in 2011–2012 from a joint initiative between the government of India and the US government.[17] This is in contrast to the US experience of civil society-led calls for open data. Nevertheless, India's open data policies did arrive into a national context where the public was making heightened calls for government accountability following the success of India's right to information (RTI) movement.[18] The RTI movement was a long-running grassroots campaign primarily by marginalized laborers and rural communities to overturn colonial-era laws limiting access to official records.[19] This mass social movement demanding transparency and accountability in government information led to the passage of India's Right to Information Act 2005, akin

16 Ministry of Electronic and Information Technology, "Government Open Data License."

17 "Open Government Data (OGD) Platform India," January 21, 2022, https://data.gov.in; Anupam Saxena, "Indian Government Launches Data.Gov.In," *Medianama*, September 4, 2012, https://www.medianama.com/2012/09/223-indian-government-launches-data-gov-in/.

18 Sumandro Chattapadhyay, "Opening government Data Through Mediation: Exploring the Roles, Practices and Strategies of Data Intermediary Organisations in India," Open Data in Developing Countries Research Network, 2014, http://hdl.handle.net/10625/60640

19 Aruna Roy and Nikhil Dey, "Fighting for the Right to Know in India," Development Dialogue.

to freedom of information legislation in many countries.[20] The RTI Act is cited in the preamble to India's open data policy, NDSAP, as a key motivation.[21]

Government-led Open Data Efforts in India: Openness Versus Control

In 2012, the NDSAP sought to make government data available for better public debate, decision-making, and to meet civil society's needs.[22] To operationalize the policy, the government launched India's open data portal (data.gov.in) to act as a platform through which users could access open datasets, modeled on the US's OGD portal (data.gov). The portal is designed to enable all ministries and public agencies of the Indian government to publish their shareable, nonsensitive datasets in an open format. The form, or "front end," of open data, therefore, is comparable to initiatives in the US and other parts of the West, requiring common standards and formats to release and integrate datasets.

Some provisions stand out regarding the conditions for data release included in India's open data policy. The policy enables government departments to decide on which datasets to share. This means each department can determine whether a dataset is shareable or non-shareable.[23] Only shareable data is contributed to the open data portal. This approach is distinct from global models which tend toward an open-by-default standard, other than where disclosure is barred by data protection or intellectual property laws.

Further, even when a department designates a dataset as shareable, the Indian policy allows access to such data to be subject to registration. Under

20 Aradhana Sharma, "State Transparency after the Neoliberal Turn: The Politics, Limits, and Paradoxes of India's Right to Information Law," *PoLAR: Political and Legal Anthropology Review* 36, no. 2 (2013): 308–25, https://doi.org/10.1111/plar.12031.

21 Government of India, "National Data Sharing and Accessibility Policy," 2012, https://data.gov.in/sites/default/files/NDSAP.pdf

22 Government of India, "National Data Sharing."

23 See paragraph 7 in "National Data Sharing."

the NDSAP access can be granted at three levels to "open" datasets: open access, registered access (after registration or authorization), or restricted access (only after specific authorization).[24] This graded access diverges from the vision for open data access to be non-discriminatory and not subject to registration. Taken together, these conditions provide a large degree of control to government departments to decide whether they release datasets, the types of data, and the type of access granted. They reflect tensions between aspirations of openness and the large degree of control over data releases by government agencies enabled by these arrangements.

Another distinct aspect of India's open data policy is its assertion of the government's "ownership" of public datasets. Even though the NDSAP recognizes that such data is gathered by public investment, its preamble frames data as an "asset." The policy repeatedly mentions its role in enabling access to "Government of India-owned data."[25] It also includes provisions that enable datasets to be priced by the "data owner" in line with government policies.[26] In 2022, the Indian government released new policy documents that propose frameworks to replace the NDSAP.[27] These proposals continue to assert ownership and control of such data by the government reflected in objectives of promoting "transparency, accountability, and ownership in Non-personal data and Datasets access [*sic*]" and the inclusion of provisions to charge "user charges/fees" toward the maintenance of open data services.[28]

These aspects indicate dual objectives within India's open data policy. While accountability and transparency are motivating factors, the policy also frames data as an asset (owned by the government) whose value is sought to be unlocked through the OGD portal. The government's assertions of ownership and control over data must be understood against the backdrop of decades of government investment in India's vast public digital

24 See paragraph 8 in "National Data Sharing."

25 See paragraphs 1.3, 3, 4, 6 in "National Data Sharing."

26 See paragraph 11 in "National Data Sharing."

27 Government of India, "India Data Accessibility and Use Policy," Ministry of Electronics and Information Technology, February 2022, https://www.meity.gov.in/writereaddata/files/India%20Data%20Accessibility%20and%20Use%20Policy.pdf; Government of India, "National Data Governance Framework Policy," Ministry of Electronics and Information Technology, May 2022, https://www.meity.gov.in/writereaddata/files/National-Data-Governance-Framework-Policy.pdf.

28 Government of India, "National Data Governance."

systems, based on a vision of the Indian government as a platform of services.[29] As Ranjit Singh's ethnography of the design team of one of India's core digital infrastructures revealed, members shared a vision in which the Indian state was cast as a database of citizen records, and the government as the arbiter in relation to such data.[30]

Such imaginaries are useful to bear in mind when unpacking the role of government in India's proposals for open data and data sharing. India's position appears to be a harbinger of things to come, given the broader trend in policy documents of governments around the world that are framing data as a strategic asset or national resource whose value must be harnessed.[31] Nevertheless, this shift indicates a departure from early non-commercial conceptions of open data. It also surfaces the difficulty of maintaining a separation between technocratic efforts to release government datasets in interoperable formats, and the political imperatives of the state as it asserts control over the same datasets.

At this juncture, it is relevant to note that the growth of the OGD platform in India has floundered in reality. The platform has suffered due to important, data-rich public agencies and departments refraining from contributing datasets to the portal, or updating past contributions.[32] Even where data is shared, its reliability and accessibility varies because departments upload PDF files instead of accessible, machine-readable formats, often with missing or incomplete data.[33] This could be interpreted as one of the consequences of the broad conditionality and delegation to government departments of the choice to share datasets in the open data policy. It could also reflect basic challenges of capacity and digital capabilities within government departments in India, or deeper issues related to the disinclination of state actors to release data.[34] These dynamics once again reveal the difficulties of

29 Ranjit Singh, "Give Me a Database and I Will Raise the Nation-State," *South Asia: Journal of South Asian Studies* 42, no. 3 (May 2019): 501–18, https://doi.org/10.1080/00856401.2019.1602810.

30 Singh, "Give Me a Database," 516.

31 Clarisse Girot, "Introduction," in *Regulation of Cross-Border Transfers of Personal Data in Asia*, ed. Clarisse Girot (Singapore: Asian Business Law Institute, 2018).

32 Thejesh G N, "Open Data in India: In a Restrictive Copyright Regime, Voluntary Organisations Pitch in to Make Data Accessible," *Engage* 55, no. 23 (2020), https://www.epw.in/engage/article/voluntary-organisations-india-counteract-states-copyright-regime-open-data.

33 Natasha Agarwal, "Unleashing the Full Potential of India's 'Open Government Data' Initiative," Ideas for India, January 25, 2016, http://www.ideasforindia.in/topics/macroeconomics/unleashing-the-full-potential-of-indias-open-government-data-initiative.html; Natasha Agarwal, "Lessons from India's (Un)Open Data," *Medium*, January 2, 2018, https://medium.com/@agarwana3/lessons-from-indias-un-open-data-on-india-s-online-visa-policy-c673469a1ad3.

34 Isha Parihar, "On the Road to Open Data: Glimpses of the Discourse in India," World Bank Blogs, February 17, 2015, https://blogs.worldbank.org/digital-development/road-open-data-glimpses-discourse-india; Neeta Verma and M. P. Gupta, "Open Government Data: Beyond Policy & Portal, a Study in Indian Context," in *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, ICEGOV '13 (New York, NY, USA: Association for Computing Machinery, 2013), 338–41, https://doi.org/10.1145/2591888.2591949.

considering the technologies of open data as operating independently from the politics and lived realities of the datafied state.

Open Data Outside the Indian State

Open data efforts in India have complex, nonlinear trajectories: one that is documented in formal policy documentation, and another that exists outside and independent of formal structures. Informal efforts to open up datasets led by software developers and independent practitioners have always existed alongside the official initiatives.[35] Even prior to the release of the NDSAP in 2012, small communities of non-government organizations and individuals had begun experimenting with aggregating, using, and resharing data and insights using datasets from government and other non-government/public sources. Coding camps on accountability and transparency in public data among networks of open data and data science enthusiasts took place before the NDSAP and continue — albeit in loose collectives, to work on issues of public interest.[36]

In recent years, clusters of volunteer-based organizations in India's tech hubs have become visible and prominent. Organizations like DataMeet, WikiData, lawresource.org, OpenStreetMap, and others were framed as "filling the gap for open data" given the floundering of the OGD portal.[37] Examples include DataMeet and DataKind Bengaluru building "data pipelines" of machine-readable data to enable greater accountability in governance by aggregating and analyzing budget data from public agencies.[38] Organizations such as Civic Data Labs and How India Lives curate data from public sources and build tools to render these datasets searchable. They are supported by web platforms, enabling users to engage with datasets through

35 Guneet Narula, "Collecting Open Data: Data Practices, Tools, Limitations and Politics," in *Lives of Data: Essays on Computational Cultures from India*, ed. Sandeep Mertia (Amsterdam, The Netherlands: Institute of Network Cultures, 2020), 108–112, https://networkcultures.org/wp-content/uploads/2020/12/LivesofData.pdf.

36 Accountability Initiative, "Code for India — Accountability & Transparency Camp," April 1, 2011, https://accountabilityindia.in/blog/code-for-india-accountability-transparency-camp/; DataMeet, "About," Data{Meet}, March 21, 2014, https://datameet.org/about/.

37 Thejesh G N, "Open Data in India."

38 Gaurav Godhwani, "Making India's Budgets Machinable," in *Lives of Data: Essays on Computational Cultures from India*, ed. Sandeep Mertia (Amsterdam, The Netherlands: Institute of Network Cultures, 2020),113-127, https://networkcultures.org/wp-content/uploads/2020/12/LivesofData.pdf.

dynamic data visualizations. Such independent platforms invest time and resources to clean, structure, and combine data from the government with other datasets available online or independently sourced.[39]

Community-led open data efforts have often responded more immediately to the needs of the public, for instance, during the COVID-19 pandemic. Several of the most effective sources of COVID-19 data came from open source collaborations and an army of volunteers who independently sourced, verified, managed, and presented data from various authorities and hospitals, due to the lack of streamlined government information.[40] The government's reasons for the failure to release effective, centralized data are complex, given that health information is often generated at the state level within India's federal structure. Researchers are beginning to unpack the reasons for these issues, including state-level disparity in the quality of COVID-19 data reporting by public bodies and issues relating to coordination and sharing of resources.[41]

The role of the official OGD portal was minimal during the pandemic. Meanwhile, despite the green shoots of community-led open data efforts being exciting, they risk becoming ad hoc or sporadic in the absence of institutionalization and consistent funding. This raises questions about the extent to which open data efforts can be truly effective in enabling accountability and civic engagement in a country like India, in the absence of genuine collaboration on open data efforts between civil society and government actors. This foregrounds the role of underlying political and social dynamics in shaping the success of technologies of open data.

39 Narula, "Collecting Open Data"; Godwhani, "Making India's Budgets Machinable."

40 Ananya Bhattacharya, "India's Best Covid Data Are Coming From Open-Source Collaboration," *Quartz*, January 28, 2023, https://qz.com/india/2118783/indias-best-covid-data-is-coming-from-open-source-collaboration; Manavi Kapur, "A COVID-19 Data Wish List For India," *Quartz*, January 17, 2022, https://qz.com/india/2113484/how-india-can-fix-its-poor-quality-covid-19-data.

41 Varun Vasudevan, Abeynaya Ganasekaran, Varsha Sankar, Siddarth A. Vasudevan, and James Zou, "Disparity in the Quality of COVID-19 Data Reporting across India," *BMC Public Health* 21, no. 1 (June 2021): 1211, https://doi.org/10.1186/s12889-021-11054-7.

Future Directions: Open Data as Technopolitical

Reflecting on India's experience with open data, it becomes evident that open data initiatives are technopolitical efforts. They reveal the difficulties of maintaining the distinction between the political aspects of opening up government data and the technocratic and technological efforts around open data that aim to release structured, machine-readable datasets.

Within such a technopolitical effort, the priorities of the datafied state matter profoundly — especially as the custodian or steward of vast troves of data gathered during its operation. Where priorities of the state shift in relation to the data it holds, they can reshape the approach or commitment to open data. A key change in the conception of government-held data emerging from the Indian case is the framing of data as an asset. This framing finds resonance in recent policy documents of other contents. The EU's Data Strategy refers to data as an "essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general."[42] The UK's recent national data strategy refers to data as a "resource" for businesses, a "vital national asset" and a "strategic asset";[43] the US Federal Data Strategy refers to data as "a strategic asset,"[44] and China has characterized big data as "a fundamental strategic resource" for the country in its 13th five-year plan (2016–2020).[45]

These statements are relevant indications of legislative intent, as seen in new proposals emerging from governments to introduce additional regulations aimed at harnessing value from sharing "non-personal data" and high-value datasets, including by structuring how they are shared across public and private actors, or across borders.[46] As these positions are fleshed out and take effect, it will invite deeper analysis of the intersection with (and influence on) the re-shaping of open data regimes.

42 European Commission, "A European Strategy for Data," 2023, https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

43 Department of Digital, Culture, Media and Sport, "National Data Strategy."

44 Office of Management and Budget, "Open Data Policy: M-13-13."

45 Communist Party of China, *The 13th Five-Year Plan for Economic and Social Development of The People's Republic of China (2016-2020)* (Beijing: Central Compilation & Translation Press, 2016), https://en.ndrc.gov.cn/policies/202105/P020210527785800103339.pdf.

46 Asmita Verma and Anjula Gurtoo, "Evaluating Global Data Policies Around Non-personal Data," Indian Institute for Science – Centre for Society and Policy, 2022, https://csp.iisc.ac.in/wp/wp-content/uploads/2022/02/Evaluating-global-data-policies-around-NPD.pdf; Olga Batura, Axel Wion, Sofia Noelle Gonzalez, J. Scott Marcus, Ilsa Godlovitch, Lukas Wiewiorra, Peter Kroon, Serpil Tas, and Nico Steffen, "The Emergence of Non-personal Data Markets," European Union, October 2023, https://www.europarl.europa.eu/RegData/etudes/STUD/2023/740098/IPOL_STU(2023)740098_EN.pdf

On the other hand, the Indian experience of open data also invites further interrogation of the expectations implicit in such technocratic initiatives regarding the citizenry in a datafied state. As some scholars have noted, shifting to the average citizen's perspective could reframe expectations of the form and conditions to which open data initiatives aspire.[47] Attending to the capacity of consumers of the data to effectively use the data could increase attention to factors such as internet access, technical requirements to use data, the usability of interfaces, language of data, etc.[48] Such an approach would focus on meeting the citizenry where they are, in contrast to the technocratic operationalization of open data portals to date — which envision tech-savvy civic-minded coders as their main audience, rather than the average citizen within the context of each country.

This essay also invites deeper reflection on open data initiatives' political role. For instance, the selective release of digital datasets may enable perceptions of greater transparency in government, regardless of whether this is actually the case. Governments may routinely release datasets and perform accountability in politically insignificant areas, even while remaining opaque in critical areas of governance.[49] Ultimately, this highlights the difficulty of separating discussions of technologies and formats of data releases from questions about the political imperatives that drive open data efforts (or resistance to them) within the datafied state.

47 Edward S. Dove, "Reflections on the Concept of Open Data," *SCRIPTed: A Journal of Law, Technology, and Society* 12, no. 2 (December 2015): 154–166, https://script-ed.org/article/reflections-on-the-concept-of-open-data/.

48 Dove, "Reflections," 159.

49 Yu and Robinson, "The New Ambiguity," 181.

COUNTERPUBLICS
COUNTERPUBLICS
COUNTERPUBLICS

BY
MATTHEW BUI and BIANCA WYLIE

COUNTERPUBLICS
COUNTERPUBLICS
COUNTERPUBLICS

COUNTERPUBLICS

By Matthew Bui and Bianca Wylie

A Counterpublic Analysis of Sidewalk Toronto

Sidewalk Toronto was an Alphabet-driven project that aimed to build up a piece of Toronto's waterfront, a small parcel of land known as Quayside (pronounced kee-side), into a testbed for digital experimentation and innovation. Although this public-private partnership (PPP) failed, it serves as a rich case study of power relations at play within tech-infused PPPs. We argue for the need for a counterpublic analysis to critically map and examine the central stakeholders involved in any PPP project and analyze the power dynamics, both formal and informal, at play. A counterpublic analysis spotlights community members and impacted parties who are given minimal opportunity to inform PPPs from the early ideation and design stage. However, these parties disparately bear the harms and risks of tech-driven initiatives. Counterpublic analyses show how communities exert power through public engagement processes.

We contend a counterpublic analysis of PPPs should ask:

- Who (that is, what communities and/or subgroups) is served by, or benefits from, the PPP?
- Are tech-driven PPPs and their processes ameliorating or re-instantiating power asymmetries?

- How do counterpublic interventions impact or increase the negotiating power of the public actor (the government/state)?
- How might a counterpublic analysis suggest interventions to increase and expand public power and agency in future projects?

PPPs are agreements between private and public actors through which private capital finances public infrastructures and initiatives in exchange for a variety of partnership benefits. Since the 1950s, projects as wide-ranging as hospitals, toll roads, bridges, water plants, and universities have been built through PPPs.[1] As the PPP Knowledge Lab explains, "There is no one widely accepted definition of public-private partnerships."[2] The construct of a typical PPP has continued to evolve, and PPP projects are beginning to include "smart" digital infrastructures more frequently. These types of digital infrastructure PPPs have repeatedly demonstrated little to no ability to productively engage in the unresolved matters of digital governance with various publics, particularly in relation to community-based concerns.[3] Despite these challenges, the appeal of smart city PPPs endure across the globe.[4]

Situating Counterpublics In Participatory Processes and Deliberation

Drawing from Gayatri Spivak's work on the subaltern[5] and Rita Felski's concept of counterpublics,[6] Nancy Fraser defines "subaltern counterpublics" as "parallel discursive arenas where members of subordinated social groups invent and circulate counterdiscourses, which in turn permit them to formulate oppositional interpretations of their identities, interests, and needs."[7] Her concept underscores how dominant discourses typically reinforce the

1 Tony Bovaird, "A Brief Intellectual History of the Public–Private Partnership Movement," *International Handbook on Public–Private Partnerships*, eds. Graeme A. Hodge, Carsten Greve, and Anthony E. Boardman (Edward Elgar Publishing, 2010), https://www.elgaronline.com/edcollchap/edcoll/9781848443563/9781848443563.00010.xml.

2 Public Private Partnership Resource Center, "What are Public Private Partnerships?" World Bank Group, last modified December 9, 2022, https://ppp.worldbank.org/public-private-partnership/overview/what-are-public-private-partnerships.

3 Germaie R. Halegoua, *The Digital City: Media and the Social Production of Place* (New York: NYU Press, 2020); Taylor Shelton, Matthew Zook, and Alan Wiig, "The 'Actually Existing Smart City,'" *Cambridge Journal of Regions, Economy, and Society* 8 (February 2014): 13–25, https://doi.org/10.1093/cjres/rsu026; Chamee Yang, "Historicizing the Smart Cities: Genealogy as a Method of Critique for Smart Urbanism," *Telematics and Informatics* 55 (December 2020): 101438, https://doi.org/10.1016/j.tele.2020.101438.

4 Burcu Baykurt and Christoph Raetzsch, "What Smartness Does in the Smart City: From Visions to Policy," *Convergence* 26, no. 4 (August 2020): 775–89, https://doi.org/10.1177/1354856520913405; Yang, "Historicizing the Smart Cities."

5 Gayatri Chakravorty Spivak, "Can the Subaltern Speak?" *Marxism and the Interpretation of Culture*, eds. Cary Nelson and Lawrence Grossberg (London: Macmillan, 1998).

6 Rita Felski, *Beyond Feminist Aesthetics: Feminist Literature and Social Change* (Cambridge: Harvard University Press, 1989).

7 Nancy Fraser, "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy," *Social Text*, no. 25/26 (1990): 56–80, https://doi.org/10.2307/466240.

status quo; it also pushes for more expansive, egalitarian visions of democracy wherein status markers are removed — that is, neutralized — within deliberation. This requires the inclusion of — and indeed, prioritizing a greater role to — subalterns, or the subordinated groups (e.g., women, workers, people of color, and queer people) minimally consulted within bourgeois, elite-driven models of deliberation.[8]

Building on this definition, we define a counterpublic analysis within tech-driven PPPs to be equally shaped by the needs and experiences of the oft-minimized (yet impacted) parties within the design, scoping, and development of interventions. Counterpublics can consist of individuals, groups, and organizations from marginalized communities, but they are generally united in, first, an experience of minimization within corporate-driven PPP processes; and relatedly, a desired goal of resisting hegemonic structures and processes in favor of more representative, and egalitarian, deliberation. Thus, our reorientation to being attendant to counterpublics within PPPs must start from the recognition of power relations and structures within democratic processes. A counterpublic analysis examines whether there are opportunities within PPPs to mitigate power asymmetries and support growth in public capacity and input. Often, this entails creating opportunities for increased public input and participation, and within it, a process that fosters democratic friction.

Sidewalk Toronto: Context and Background

Sidewalk Toronto was described as an opportunity for local innovation, designed to address pressing urban problems and premised on values of sustainability and inclusion. The project was a direct collaboration between a

8 Fraser, "Rethinking the Public Sphere," 70.

public entity, Waterfront Toronto, representative of all three orders of the Canadian government, and a private company, Sidewalk Labs, an Alphabet subsidiary. Sidewalk Labs was the successful winner of a request for proposals to create and fund a plan to develop a parcel of valuable real estate on Toronto's waterfront, named Quayside.

Beyond their plans for real estate development, the development of tall-timber building construction, modular housing ideas, and autonomous vehicles, Sidewalk Labs' initial proposal also included ideas to transform Quayside into a testbed for technologies by merging the city's physical and digital layers. These included deploying ubiquitous connectivity, installing data sensors to monitor air quality, noise levels, automobile and pedestrian traffic, and weather; combining census data, open data, and Google data to power simulation models and portals; and building high-tech infrastructure, such as autonomous sanitation systems and mail delivery.[9]

Overall, the Sidewalk Toronto PPP aimed to leverage Sidewalk Labs' corporate ties and capital to improve and invest in Toronto's infrastructure. This positioned both Alphabet and the City of Toronto as global leaders in high-tech urban innovation; and expanded Alphabet's foray into real estate development and traditional capital infrastructure financing (e.g., a streetcar line). Amid COVID-19 budgetary concerns, the partnership was terminated by Sidewalk Labs in May 2020.

9 Sidewalk Labs, "Sidewalk Labs Vision," October 17, 2017, https://storage.googleapis.com/sidewalk-labs-com-assets/Sidewalk_Labs_Vision_Sections_of_RFP_Submission_7ad06759b5/Sidewalk_Labs_Vision_Sections_of_RFP_Submission_7ad06759b5.pdf.

Situating Counterpublic Power in Relation to the Public

If we understand the public in public-private partnerships to strictly mean the government as the representative of the public, one could argue that the general public, writ large, has little to no role to play in negotiating PPP contracts. Historically, this has been the case. Even when broader conceptions of publics beyond the government are involved in deliberative processes, Fraser contends that elite-driven models reify the historic and structural subordination of counterpublics: namely, they render members of various (marginalized) social groups as minimally legible within conceptions of publics, thereby warranting this alternative category.[10]

Despite the design of processes that typically seek to exclude and remove their power, counterpublics' self-directed participation in the Toronto process expanded and grew the negotiating power of the public actor (Waterfront Toronto) within the partnership. By contesting the project, various counterpublics created additional room and pressure for Waterfront Toronto to demand improved terms for the deal. The final stages of negotiation, by which Waterfront Toronto made increasingly beneficial public value demands, are one of the inputs that led to the demise of the project by impacting the projected profitability and project scope.

10 Fraser, "Rethinking the Public Sphere."

Which Publics Have a Say in a Public-Private Partnership?

The public participation component of Sidewalk Toronto would more properly be understood as an extractive model designed to inform Sidewalk Labs' product development. In the framing of the public engagement processes, as created by both Waterfront Toronto and Sidewalk Labs, saying no to the project was not an option. Waterfront Toronto and Sidewalk Labs wanted to hear from the types of publics enthusiastic about helping shape the plan, who generally thought the partnership and approach was a good idea. This inherently assumed general consent from the entire city's population regarding the proposed partnership and development plans. *Saying no* to the project wasn't on the table for the public, only helping with the how.

This core framing problem underscores important critiques, previously mentioned, regarding which publics constitute the purview of PPPs, including the flattening of multiple publics into one public and portraying public and private benefits as equally tiered. As the project unfurled from real estate development and land valuation to economic development and intellectual property, from transportation automation to neighborhood technology infrastructures there were some community members or civic institutions who enthusiastically favored the proposals, and some community members and institutions who deeply opposed them.

Issue by issue, the breadth and complexity of the proposed project surfaced critiques from multiple perspectives. The negotiation process of the deal was subject to an unusual amount of public oversight for a PPP because a range of publics and counterpublics refused the corporate capture that

attempted to set the terms and boundaries of public engagement. Their refusal forced public leaders in charge of negotiations to ensure that the terms of the deal would stand up to deep public scrutiny. This was a display of public — and counterpublic — power, despite both the public and private actors failing to frame the process in support of this kind of conflict and refusal.

A key appeal in the tech-infused PPP narrative is that the technology sector, as the well-capitalized and creative force, can fill in the gaps for governments, particularly amid times of austerity.[11] This was, and continues to be, a major vulnerability in Toronto, one common in many cities, where residents are frustrated by aging infrastructures, lack of affordable housing, and cost-of-living increases. The selling points used to pitch the project to the general public were less about technology and more about quality of life.[12] In short: "Your government can't do what's needed. Tech companies can."

Alarmingly, in Sidewalk Toronto as well as other smart city projects, digital technologies are named as key tools for how to improve quality of life. Tech-infused PPPs often gesture toward the need for increased data collection and use to improve public spaces and services. Yet as Chris Gilliard and David Golumbia underscore, it is the privileged (i.e., wealthier and white) communities that can more easily opt out of these technologies and avoid consequences, compared to the minoritized (i.e., poor, immigrant, and/or BIPOC) communities more likely to bear their risks and harms.[13] Gilliard and Golumbia call for reassessing who benefits from tech-driven interventions. A counterpublic analysis asserts the importance of counterpublics in reimagining interventions that address pressing, community-relevant problems while also allowing for their continued engagement and input as projects evolve, foregrounding this unequal distribution of negative impacts and privileges.[14]

11 Baykurt and Raetzsch, "What Smartness Does in the Smart City."

12 Bianca Wylie, "Debrief on Sidewalk Toronto Public Meeting #1— Evasive on Data Products, No Answer on Data…," *Medium*, March 28, 2018, https://biancawylie.medium.com/debrief-on-sidewalk-toronto-public-meeting-1-evasive-on-data-products-no-answer-on-data-a9f551535dcd.

13 Chris Gilliard and David Golumbia, "Luxury Surveillance," *Real Life Mag*, July 6, 2021, https://reallifemag.com/luxury-surveillance/.

14 Gilliard and Golumbia, "Luxury Surveillance."

The Interplay of Publics, Counterpublics, and Representative Power

Sidewalk Labs continuously sought to influence and win the support of the City of Toronto's civic elites. They organized events and participation models to court members of various publics that included local neighborhood associations, former elected officials, members of the political class, leaders of nonprofit organizations, the technology start-up community, volunteer organizations, and so forth. They did this mostly in partnership with Waterfront Toronto, and in some cases independently.[15]

Across many sectors of civic life in Toronto, there were people and groups that welcomed the project wholesale, that would only conditionally support it, and others that rejected the idea and wanted to refuse the project entirely. Below is a small sample of the kinds of topical tensions, and related inter-community frictions, that were in effect between some of the representatives of various publics and counterpublics implicated in the Sidewalk Toronto process.

Affordable housing. ACORN, an affordable housing advocacy organization, was relentless in challenging both the public and private partners for designating too few units of affordable housing on public lands during a housing crisis. The concerns that these housing advocates brought into the conversation were minimized in relation to conversations about technology. At the first public organizing meeting held by BlockSidewalk, the full room in attendance (100+ people) agreed on the need to prioritize land use for affordable housing. This public demand was immensely difficult to keep in the conversation during the project's duration. The profile of affordable housing had to fit within the conservative vision regarding the number of units to

15 Josh O'Kane, "Sidewalk Labs Forming Separate Advisory Panel for Toronto Smart-City Project," *The Globe and Mail*, October 9, 2018, https://www.theglobeandmail.com/business/article-sidewalk-labs-forming-separate-advisory-panel-for-toronto-smart-city/.

be made available in the development proposed by both Waterfront Toronto and Sidewalk Labs. As with everything else in the project, the innovations proposed regarding housing construction and the attendant potential market for these goods overshadowed the actual number of affordable housing units that the project would create.

Privacy. An alternative model for data management — a civic data trust — was proposed during the project. The former Ontario Privacy Commissioner came out against the idea of a community stewardship model for data collection and use, as did the sitting Ontario Privacy Commissioner.[16] While privacy professionals have long upheld the privacy rights of Canadians, most of them were not interested in considering a model that might take a more expansive look at how various publics and counterpublics could potentially organize around data governance. Here again, counterpublics with different concerns extending beyond (technical notions of) privacy were not given the same status and stature in conversations as those held by the privacy establishment. This conservative approach also played into Sidewalk Labs' framing: if the project was privacy-preserving, then it should be a go. Such framing forestalled discussions about the full-fledged privatization of local governance in Toronto, an issue upstream of (and larger than) privacy.

Academic research. Many universities in Toronto signed on to Sidewalk Labs' grant-funded projects, and in doing so appeared to prioritize their institutional desires to be part of something innovative and visible in the press over the needs and concerns of some of the residents that these public institutions are implicated in representing. Universities lent significant credibility to the project by highlighting their importance in the innovation economy, but without accountability to the counterpublics and critics who held starkly

16 Donovan Vincent, "Sidewalk Labs' Urban Data Trust Is 'Problematic,' Says Ontario Privacy Commissioner," *Toronto Star*, September 26, 2019, https://www.thestar.com/news/gta/sidewalk-labs-urban-data-trust-is-problematic-says-ontario-privacy-commissioner/article_ae44fec0-2180-58f3-8799-196a034707ce.html.

different views of the project's fundamental impacts. Public universities, like state actors, are under fiscal and political pressures to take part in, and support, the innovation economy, rather than concerned counterpublics.

Urban planning. Urban planning professionals, particularly those who are registered, have a duty to the public, and for some registered professionals, a code of ethics that commits them to doing work in the public interest. Some professional planners worked with the private partner to advance Sidewalk Labs' interests, sometimes motivated by frustration with local government and its lack of interest in trying new things. Other urban planners critiqued the project and worked on the side of the counterpublics: they worked both within and from outside government to challenge the dominant approach and upsides of innovation that were marketed to the city.

Economic development. The Canadian business community is an exception to the type of group that usually comprises a counterpublic, but in this case study, it must be mentioned that the geopolitics of their dissent about the project bore significant weight in the political discourse. Some of the local tech startups were excited to take part in the project. The Toronto Region Board of Trade was a vocal supporter; others, such as the Canadian Council of Innovators, were in steadfast and vocal opposition.

These are but several examples. The list is non-exhaustive. The intent is to reflect on the wide range of smaller and less visible topical frictions — and implicated counterpublics — that were engaged in the conversation, and how their concerns were subsumed beneath various public interest actors. The negation of their concerns was especially pronounced when these counterpublics refused to support techno-solutionism as a model and general approach to city building. Counterpublics would have borne the largest risks of the project.[17] These harms included the obfuscated risks of

17 Adwoa Afful, "Toronto Can't Be a Futuristic City," *Bitch Media*, January 15, 2019.

privatization, such as turning over public design, maintenance, and oversight of digital public infrastructures to private entities, worsening an already opaque process for accountability and redress.

Practical Lessons From Sidewalk Toronto

Procurement as a site for counterpublic advocacy in PPPs. In the context of Sidewalk Toronto, it was the state — through three levels of government — that had, and has, a democratic duty to all publics and counterpublics. In their failure to own up to this role, they enabled a private actor, one with relationships within a consumer context and not a democratic one, to wield influence that was not theirs. As Bianca Wylie elaborates, this negligence to support the counterpublics — to whom the state is accountable — was designed into the process right from the start via the state-created and designed request for proposal.[18] That is, the request for proposal process serves as a prime example of how elite-driven models for deliberation reify the subordination of counterpublics.

Future advocacy efforts should consider the procurement phase of any digital infrastructure project as a potential area for engagement, refusal, and resistance, particularly of counterpublics. This includes participating in proactive disclosure advocacy: requiring governments to communicate with residents about potential digital infrastructure projects prior to writing tendering documents. Another policy advocacy opportunity is seeking commitments from governments to mandate engagement on the proposed tendering process for PPP projects of a certain size or type, actively seeking to move beyond the traditional boundaries of public participation that focus on privileged groups.

18 Bianca Wylie, "In Toronto, Google's Attempt to Privatize Government Fails — For Now," *Boston Review*, May 13, 2020, https://www.bostonreview.net/articles/bianca-wylie-sidewalk-labs-toronto/.

From ownership to stewardship: Designing and supporting community self-governance of digital infrastructures. Mandating the ongoing participation of counterpublics in the governance of new digital infrastructures is a tactical opportunity to shift power. For one, counterpublics, as extensions of larger communities, can advocate for the creation of ongoing stewardship models in the governance of neighborhood technologies, a distinct departure from prevailing top-down models of tech ownership and control. By creating new self-governance models and advocating for public funding to support their operations, residents can build up a more persistent approach to both governing and refusing the use of technology. Counterpublics can define acceptable norms and create friction in cases where technology must be refused, removed, or put on hold. By setting up ongoing oversight with public participation, private companies will also have to grapple with what it means to consistently engage with (counter)publics. As a result, public bids will likely require more flexible, transparent, and adaptable approaches to product development and maintenance. In this manner, the state, through engagement with various counterpublics and publics, can leverage public power and funds to reshape the public technology market. This could include the use of mandatory technology standards in procurement.

Designing self-governance considerations as a requirement for bidders shifts the public mindset from accepting what the market has to offer to asserting what it needs from vendors. Modes of increased participatory governance or self-governance — wherein community members play more active roles in shaping and determining interventions and outcomes — are seen in community land trusts,[19] data and digital infrastructure trusts,[20] digital justice principles,[21] commons models,[22] civic co-design models,[23] and public and digital realm stewardship.[24] We contend this approach could reorient interventions in favorable ways toward increased participation designed by

19 "The Community Land Trust Model and Movement," https://groundedsolutions.org/tools-for-success/resource-library/community-land-trust-model-and-movement; Bianca Wylie and McDonald Sean Martin, "What Is a Data Trust?" Centre for International Governance Innovation, October 9, 2018, https://www.cigionline.org/articles/what-data-trust/.

20 Wylie and McDonald, "What Is a Data Trust?"; Sean Martin McDonald, "Reclaiming Data Trusts," Centre for International Governance Innovation, March 5, 2019, https://www.cigionline.org/articles/reclaiming-data-trusts/; Sean Martin McDonald, "Civic Data Trusts," Some-thoughts.org, Accessed November 20, 2023, https://some-thoughts.org/mcdonald.html.

21 Nasma Ahmed, "Digital Justice Principles," 2019, https://www.some-thoughts.org/ahmed.html.

22 Kristin Hayes, interview with Erik Nordman, Resources Radio, podcast audio, March 8, 2022, https://www.resources.org/resources-radio/managing-the-commons-insights-from-elinor-ostrom-with-erik-nordman/.

23 Sheila R. Foster and Christian Iaione, *Co-Cities: Innovative Transitions toward Just and Self-Sustaining Communities* (Cambridge: MIT Press, 2022).

24 Bianca Wylie and Zahra Ebrahim, "Shared Governance: A Democratic Future for Public Spaces," *Azure Magazine*, February 3, 2021, https://www.azuremagazine.com/article/bianca-wylie-zahra-ebrahim-shared-governance-public-space/.

and for counterpublics, especially when combined with interventions that target procurement. Inserting and requiring self-governance models to be part of any PPP project opens more possibilities for ongoing participation and adaptation.

Conclusion

Public-private partnerships involving digital infrastructures, such as Sidewalk Toronto, emphasize data and data-driven technologies in ways that threaten to replace, weaken, or delegitimize democracy. As researchers and advocates, we propose the development of case studies and critical frameworks that proffer counterpublic analyses to foster this reckoning with the power relations — and differentials — laden within PPPs. It allows us to focus narrowly on the specific complexities, tensions, and conflicts present within projects, even within the flattened category of public.

Moving beyond the unhelpful flattening of power relations within notions of universal public interest, (a problem well-known in urban planning circles), a counterpublic analysis underscores important questions about power and inequality, which are overlooked within simpler notions of the "public" within PPPs.[25]

Like Fraser, we do not anticipate the process to be simple and seamless, nor do we claim that all counterpublics are well intentioned and siding with the public interest over corporate expansion.[26] Yet, an emphasis on counterpublics, when coupled with increased participatory governance, enables multiple rounds of deliberation and facilitation in order to ensure increased influence throughout all stages of digital infrastructure pre-building, design, and maintenance. Aligning ongoing public participation with the full life

25 See also Maria Filippelli, "Public Interest Technology" and Anne L. Washington and Joanna Cheung, "Public Interest," *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

26 Fraser, "Rethinking the Public Sphere."

cycle of technology is one of the most opportune approaches to being able to operate with a true and defensible social license. Governments that are willing to create, support, and fund community self-governance arrangements for digital infrastructures can ensure increased public guidance and oversight throughout the PPP process, rather than treating engagement (and impacted parties) as an afterthought.

Admittedly, the proposed intervention is only a possible beginning and not the end of an exploration of where and how we can reimagine counterpublic engagement for the broader — and democratic — good. Moreover, based on a North American example premised upon democratic principles, this model will manifest differently within other social, cultural, and political contexts and models for tech ownership, stewardship, and financing. In the end, we draw attention to how PPPs disparately distribute harms and risks, and thus the need to shift power relations in a more nuanced, equitable, and impacted-first manner. We call attention to the leakage of public power and the alarming implications if such corporate power is not checked. We also draw attention to the importance, and wide range of complex views, among counterpublics that should be earnestly considered to mitigate the disparate risks and harms of tech-infused PPPs. Through engaging counterpublics, tech-driven PPPs can preserve and enable, rather than forestall, democracy. We also must resist the replacement of democratic institutions by technological processes that remove and reduce the input and engagement of counterpublics. The alternative is increasingly anti-democratic technocratic systems of governance, owned and managed by private interests, that seek to continue the ongoing blurring of the line between resident and consumer.

COUNTERDATA
COUNTERDATA
COUNTERDATA

BY

SEYI OLOJO

COUNTERDATA
COUNTERDATA
COUNTERDATA

COUNTERDATA

By Seyi Olojo

Introduction

Communities, activists, and organizers have long used data to narrate their everyday lives. Data, often situated as a tool for truth-telling, is at the foundation of empirical research and investigative reporting, but data has also long been a tool for storytelling and grassroots resistance. For historically marginalized communities, one way of controlling the narratives about them is through data activism.[1] This includes collecting their own data to either contest state-led projects of legibility or to collect missing data.[2] *Counterdata* is data that is collected to contest a dominant institution or ideology. While the practice of counterdata emerged from communities, it has been conceptualized and theorized within geographic scholarship and has gradually found its way into emerging research fields, such as critical data studies and human-computer interaction.

Counterdata can be defined as data produced as a means for enabling disadvantaged communities to reclaim political power. Key components of counterdata include (1) the correction of misrepresentative data, (2) the control of data collection and production, and (3) the strategic use of data to benefit the political and social emancipation of communities. It's important to note that counterdata is also reflective of a relationship between the state and its subjects, therefore it is inherently reactionary. Later, I'll discuss the possibilities of data activism beyond reactionary purposes. Counterdata is

1 Stefania Milan and Miren Gutierrez, "Technopolitics in the age of Big Data," *Networks, Movements and Technopolitics in Latin America: Critical Analysis and Current Challenges,* eds. Francisco Sierra Caballero and Tommaso Gravante (Switzerland: Palgrave Macmillan, 2018), 95–109, and Catherine D'Ignazio, "Chapter 1 — A Short Genealogy of Feminicide and Data Activism," in *Counting Feminicide: Data Feminism in Action*, https://mitpressonpubpub.mitpress.mit.edu/pub/cf-chap1.

2 Alessandra Jungs de Almeida, Lauren Klein, and Catherine D'Ignazio, "Missing Data," *Keywords of the Datafied State,* eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

not simply limited to the question of who collects the data. This term is rooted in practice and aims to describe the work that activists do in using data to realize a collective consciousness about a political issue. As a result, not only does counterdata reflect a history of data-driven grassroots methods of resistance; but it also names a shift in the way data is understood and conceptualized within statistical reporting.

Counterdata only begins to describe the ways communities and organizations subvert official datasets, given that contestation of official data has long been a tactic used by grassroots organizations. At present, state control is both mobilized and reinforced through data-driven technologies. In the United States, communities of color are continuously surveilled through federally funded data collection efforts and historically, this surveillance has led to significant structural discrimination. Well-documented examples include gerrymandering to promote voter suppression, the over policing of predominantly African American neighborhoods, and unequal distribution of federal funding for public schools. Additionally, technological failures caused by the incorrect use of data continue to harm communities that have historically been victimized by bureaucratic institutions. For example, a 2019 *Washington Post* article summarized the findings of a study conducted by The National Institute of Standards and Technology (NIST) that found a statistically significantly high rate of false positives within the racially biased facial recognition technology used by law enforcement.[3] Ultimately, the justification for this surveillance is embedded within state projects for resource allocation and management. As impacted communities realize the explanatory power of data, numerous grassroots initiatives have collected counterdata to contest numbers or official statistics produced by institutions. This keyword essay attempts to summarize the emergence of the term within critical data studies while specifically historicizing it within the

3 Drew Harwell, "Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use," *Washington Post* (2019).

context of radical Black methodologies and grassroots organizing, which persists today. In addition to this keyword entry, Massaro et al.[4] provide a case study analysis of counterdata production against recidivism within a Pennsylvania correctional facility.

Black Methods: Rewriting Narratives With Data

Ida B. Wells' investigative work provides an early and influential example of counterdata. Wells, an investigative journalist, and researcher, collected statistics on lynching in the late 19th century. She famously collected and curated her own datasets in an effort to show the public the alarming and often unjustified rate at which Black men were being lynched in America. Her collection of data disaggregated by race dared to fill the missing information on the landscape of lynching in the United States.[5] In this process, she was able to debunk myths surrounding the nature of Black life that were often used to falsely justify killings.

Wells was not only contesting the state's misreporting of lynching deaths, nor the pervasively dehumanizing narratives of Black life; she was also responding to a history of quantification by acknowledging its deep racist and eugenic origins.[6] Wells' anti-lynching pamphlets became famous for changing the narrative of lynching, which at the time was considered a justified punishment for people charged with rape. Black men overrepresented the number of men who died annually from lynching with alleged charges of rape or attempted rape. Famously, the *Chicago Tribune* reported lynching statistics every year, containing information on the alleged offense, and the race of the accused. The *Tribune's* report failed to highlight that the lynchings were racially motivated, rather than the result of accusations of rape.

4 Jungs de Almeida, Klein, and D'Ignazio, "Missing Data."

5 Ida B. Wells-Barnett, "Lynching and the Excuse for It," *The Independent* 53, no. 2737 (1901): 1133–1136.

6 Anne M. Brubaker, "Who Counts? Urgent Lessons from Ida B. Wells's Radical Statistics," *American Quarterly* 74, no. 2 (2022): 265–93.

Wells knew that the *Tribune's* data was incorrect. They failed to aggregate the data by race, which would have illustrated the disproportionate amount of Black people that made up a majority of lynching deaths. In turn, Wells published *The Red Record*, which used her own empirical investigation and the *Tribune's* data to correct its shortcomings. She provided percentages of Black people killed in proportion to the overall reported to be killed within the *Tribune*. She also specifically provided aggregates according to the reported offense. This extra step of *descriptive statistics*, adding analysis to data sources from white media outlets, was used by Wells to correct a false narrative through a quantitative approach, understood at the time to be irrefutable.

Another foundational example can be found in the work of the pioneering sociologist, Dr. W. E. B. DuBois. Insistent on presenting a narrative of Black life that empirically captured the nuances of Black communities, DuBois was in constant search of data. As a result, the data visualizations he created for the 1900 Paris exhibition were the world's very first infographics on Black life.[7] Additionally, DuBois's *Philadelphia Negro*[8] was amongst the first in-depth sociological studies of African-American life in an urban environment. His survey data and analysis eventually led to the release of a special report that interrogated the then-normative and limiting theories of race in America. Both Wells and DuBois understood that collecting data, often missing data, to contest official statistics would open up new possibilities for how society could come to know Black life. Perhaps the racist ideologies promoting a monolithic representation of Blackness could be corrected. Nonetheless, the struggle to correct false narratives persists, but correction serves as one of the primary motives behind the counter production of data. Currently, data remains at the center of liberation and political mobilization for Black communities. In addition to the task of correcting false statistical reporting, community-based organizations have prioritized fighting for the control of their data.

7 Witney Battle-Baptiste and Britt Rusert, *W.E.B. DuBois's Data Portraits: Visualizing Black America,* (New York: Princeton Architectural Press, 2018).

8 W. E. B. DuBois, *The Philadelphia Negro: A Social Study* (Philadelphia: University of Pennsylvania Press, 1996).

Today, organizations like Data for Black Lives, COVID Black, and Campaign Zero exist as the custodians of the most reliable databases for COVID-19 and police violence reporting for Black communities. Data for Black Lives, which rose in popularity during the pandemic, was one of the first organizations to specifically track COVID-19 data amongst Black communities. Created in reaction to the lack of accurate reporting on the impacts of COVID-19 on Black communities, Data for Black Lives created a network of scientists at various city hubs in the country to collect and report on local COVID-19 data. Another organization that uses data to correct narratives on Black life is Campaign Zero. Its Police Data Transparency Index reports on police activity and police misconduct across each state. Additionally, COVID Black, an organization that was created during the pandemic, gathers and publishes data on Black health. In partnership with academic institutions, healthcare organizations, and Black communities, COVID Black produces Black health data that helps organizations report more accurate statistics on the state of Black health. These organizations have strategically situated themselves as reputable data sources that counter the authority of institutions. By doing this, they maintain data ownership. Reminiscent of Well's original vision of better subverting racist justifications for the lynchings, these organizations maintain and produce data sources to hold policymakers accountable for the epidemic that is premature death within Black communities.

Counter-mapping

The conceptual work of the prefix counter — has shown up in many disciplines. Counterdata has clear relationships to concepts like counterpublics, or even theories of the counterculture, both of which discuss how groups

organize in opposition to central power. However, I believe a key aspect of counterdata — specifically the role of data — can be understood by examining the history of counter-mapping. Counter-mapping is a concept developed in critical geography, reflecting on the historical role of maps in the power of the state. Critical geographers Craig Dalton and Jim Thatcher coined the term "counterdata actions" in reflection of previous scholarship in counter-mapping.[9] They go on to cite Nancy Peluso's seminal article describing the Indigenous processes of forest mapping in Kalimantan, Indonesia as divergent from dominant mapping practices.[10] Here, we first see counter-mapping defined. Peluso recalls how the Indonesian state's techniques for rendering communities within limited logistics of representation, so that customary natural resource allocation practices may remain intact. Counter-mapping, in her essay, makes visible the fact that mapping has always been highly political. We see this in the idea of *terra nullius*, uninhibited land, that promoted Western frontierism. New maps were constructed to affirm the privatization of indigenous lands, therefore creating the grounds for counter-mapping to take deliberate departures away from dominant institutions and methods for legibility.

Even though maps have historically been used for capital accumulation and to promote state legibility, critical geographers are imagining new ways of mapmaking. Within situated geographic practices, we've seen a break from traditional modulations of data structures to embrace technical relations that work better for the communities the data aims to serve.[11] Examples of this can be seen within scholarship depicting counter-mapping practices and its bottom-up approach to resisting exploitative relationships between people and capital accumulation.[12] Additionally, counter-mapping helps situate a theory of counterdata which is reminiscent of Indigenous histories of resisting dispossession. They are explicit in asserting

9 Craig M Dalton, Linnet Taylor, and Jim Thatcher, "Critical Data Studies: A Dialog on Data and Space," *Big Data & Society* 3, no. 1 (June 2016): https://doi.org/10.1177/2053951716648346.

10 Nancy Lee Peluso, "Whose Woods Are These? Counter-Mapping Forest Territories in Kalimantan, Indonesia," *Antipode* 27, no. 4 (1995): 383–406, https://doi.org/10.1111/j.1467-8330.1995.tb00286.x.

11 Clancy Wilmott, "Small moments in spatial big data: Calculability, authority and interoperability in everyday mobile mapping," *Big Data & Society* 3, no. 2 (2016), https://doi.org/10.1177/2053951716661364.

12 Craig M. Dalton and Tim Stallmann, "Counter-Mapping Data Science," *Canadian Geographies / Géographies Canadiennes* 62, no. 1 (2018): 93–101, https://doi.org/10.1111/cag.12398.

that counterdata actions are already happening, making clear that this concept was developed from grassroots approaches to subvert state control. Furthermore, counterdata can be produced for a multiplicity of reasons, but it can be seen as data practices that are in accord with the political and epistemological visions of the communities they represent. Even though counterdata production adheres to the logic of quantification, a long-preferred method of empiricism, communities are able to prioritize which numbers are most meaningful to them. Counterdata centers marginalized communities as the main stakeholders with the technical skills to create artifacts that tell narratives that are most authentic to their lived experience.

Counterdata: Defined by Case Studies

Even though counterdata can be connected back to other theories of data activism, it is important to note that counterdata production has always been a dynamic global grassroots practice. Counterdata has always been produced to return agency to oppressed communities. Given its origins in grassroots organizing and social movements, most scholarly writing on counterdata, much of it emerging in the past decade, theorizes the concept using case studies. This notably includes work following femicide data collection in Latin America[13] and data collection surrounding the epidemic of missing and murdered indigenous women and girls. These examples of data collection as a response to misogyny-driven killings and abuse highlight the important role that counterdata plays in disrupting the power relations that missing data[14] maintains. In her work on the mapping of femicide accounts in Uruguay, activist-scholar Helena Suárez Val[15] describes the affordances of collecting this data, "In this sense, digital records and cartographies of feminicide are a form of research-creation where data about violence becomes

13 Harini Suresh, Rajiv Movva, Amelia Lee Dogan, Rahul Bhargava, Isadora Cruxen, Angeles Martinez Cuba, Guilia Taurino, Wonyoung So, and Catherine D'Ignazio, "Towards Intersectional Feminist and Participatory ML: A Case Study in Supporting Feminicide Counterdata Collection," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22 (New York, NY, USA: Association for Computing Machinery, 2022), 667–78, https://doi.org/10.1145/3531146.3533132; Catherine D'Ignazio, Helena Suárez Val, Silvana Fumega, Harini Suresh, Isadora Cruxên, Wonyoung So, Ángeles Martinez, Mariel García-Montes, "Feminicide & Machine Learning: Detecting Gender-based Violence to Strengthen Civil Sector Activism," Mechanism Design for Social Good Workshop, (August 2020). http://www.kanarinka.com/wp-content/uploads/2021/01/DIgnazio-et-al.-2020-Feminicide-Machine-Learning-Detecting-Gender-ba.pdf.; Helena Suárez Val, "Datos discordantes. Información pública sobre femicidio en Uruguay," *Mundos Plurales-Revista Latinoamericana De Políticas Y Acción Pública* 7, no. 1 (2020): 53–78.; Helena Suárez Val, "Datos discordantes. Información pública sobre femicidio en Uruguay,." *Mundos Plurales-Revista Latinoamericana De Políticas Y Acción Pública* 7, no. 1 (2020): 53–78; Sarah Meagan Upton, "Moving Beyond Awareness: Ni Una Más and Approaches to the Problem of Femicide in Ciudad Juárez," (MA Thesis, Georgetown University, 2010).

14 Jungs de Almeida, Klein, and D'Ignazio, "Missing Data."

15 Helena Suárez Val, "Affect Amplifiers: Feminist Activists and Digital Cartographies of Feminicide," in *Networked Feminisms: Activist Assemblies and Digital Practices,* eds. Shana Macdonald, Brianna I. Wiens, Michelle MacArthur, and Milena Radzikowska (Lanham: Lexington Books, 2021): 163–187.

public displays of feminist activists' emotional and affective — and political — responses to feminicide."[16]

Suárez Val has been tracking feminicide in Uruguay since 2015. In her essay, "Affect Amplifiers: Feminist Activist and Digital Cartographies of Feminicide," she argues that feminist rearticulations of feminicide data create the conditions for social change. The act of accounting for these deaths, through recordings, mappings, and visualizations, invokes feminist affective politics that effectively contest the rampant nature of feminicide. Val situates her work in the feminist affective cultures of Uruguay, recalling local histories of women advocating against violence against women in the 90s. The maps that Val presents in this paper center the emotions of feminist activists on the front lines of reporting and data collection. Data collection isn't simply the act of sourcing data, it holds the weight of the lives lost. Val describes one map in particular where data points were illustrated through the use of tear shapes. The atrocity of these deaths is communicated beyond the epistemological boundaries of numbers.

Another example of counterdata can be found in the work of Amanda Meng and Carl DiSalvo.[17] In partnership with the Westside Atlanta Land Trust (WALT), they observed how this community was able to collect its own data to advance its advocacy for a community land trust. The official data, a survey of the built environment, was conducted in 2014 to assess property vacancies in the English Avenue and Vine City neighborhoods. The survey data was supposed to help local organizations combat gentrification, but after reviewing the survey data, they quickly determined that the data significantly undercounted abandoned property. WALT, one of many groups advocating for the development of a community land trust in Fulton County, was able to spot this issue because of their own lived experience of their

16 Suárez Val, "Affect Amplifiers," 1.

17 Amanda Meng and Carl DiSalvo, "Grassroots Resource Mobilization Through Counter-data Action," *Big Data & Society* 5, no. 2 (2018): https://doi.org/10.1177/2053951718796862.

neighborhood. This local knowledge propelled a collective resistance to the official data. For them, data collection not only created a new artifact but encouraged a collective consciousness amongst the people. Their data strategically contained information, such as back tax amounts for each property. Properties that had back taxes were flagged as potential sites for the community land trust. The organized resistance against the county's original survey created an opportunity for community members to think about datasets that would be of better use to their communities and that would effectively subvert state power.

Morgan Currie, Britt S. Paris, Irene Pasquetto, and Jennifer Pierre[18] illustrate an example of counterdata production in the reporting of Police Officer Involved Homicides (POIHs) in Los Angeles County. Drawing from Dalton and Thatcher's seminal paper, Currie et al. provide another case study approach to defining the term. They define counterdata actions as "acts of resistance to politically dominant datasets." In this case, we see data produced in ways that center communities as the primary designers of the data. In the absence of official POIH data, Currie et al. observe how communities reinterpret previously existing datasets that fail to accurately capture the landscape of law enforcement homicides in their communities. They describe findings from a community hackathon — showing that participants were able to leverage their technical skills in spotting inaccuracies in federal datasets while also creatively thinking of ways to index and process community-derived datasets. In addition to making community-derived data more accessible, participants resisted the boundaries of quantitative data collection. They researched the social media profiles of victims to shed light on the lives they lived and the communities that memorialized them. Here, we see counterdata as not just a quantitative approach to official data, but as sites for qualitative and interpretivist methodologies. Not only are authors

18 Morgan Currie, Britt S. Paris, Irene Pasquetto, and Jennifer Pierre, "The Conundrum of Police Officer-involved Homicides: Counter-data in Los Angeles County," *Big Data & Society* 3, no. 2 (December 2016), https://doi.org/10.1177/2053951716663566.

dealing with a missing data issue (no official database of POIH data) the collection of their counterdata challenges policing narratives in LA. In this case, we see that counterdata actions allow for access to information. This includes challenging the inadequate ways in which data has been made available. Qualitative data collection often exists as a form of counterdata within research environments that fetishize quantitative methods. The above examples assert the importance of including local knowledge and narratives in datasets that are able to better represent communities. Qualitative methodologies can often shed light on the incomplete and incorrectly labeled information contained within official datasets.

Limitations and Conclusion

Counterdata actions can be instructive in how researchers think about data as a representative tool. State and institutional powers have almost exclusively reduced data into ledgers that are made to produce a surveillant order to everyday life.[19] The history of statistics is embedded within this goal of building the nation-state[20] and promoting eugenic ideals.[21] Counterdata has consistently purported that qualitative approaches to accounting for people not only produce more accurate data but provide opportunities to contest subjectivity. Inherently resistant to the objectives of scale and generalizability for hegemonic use, might it be possible to see qualitative data collection and ethnographic practices as political resistance?

Counterdata is corrective, which means that it will always be produced in response to incorrect official data. It can illuminate hierarchies of power embedded within datasets and furthermore, politicize previously depoliticized datasets. This always situates the hegemonic institutions as political

19 James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 2020).

20 Mara Loveman, *"Nation-state Building, 'Race,' and the Production of Official Statistics: Brazil in Comparative Perspective*," (Ph.D. Diss., University of California, Los Angeles, 2001).

21 Ian Hacking, "Trial by Number; Karl Pearson's Chi-Square Test ....," *Science* '84, no. 5 (November 1984): 69–71.

actors and data as tools for resistance. But what can data outside the context of resistance afford communities? Indigenous data practices have often pushed back against the limitations of counterdata in favor of data practices that primarily serve their own communities.[22] Nonetheless, the rhetoric around counterdata is still constrained by the construct of the statistical dataset — an inherently quantitative tool.

Finally, this essay asserts that the work of defining counterdata originated from grassroots organizing and activism. This term evolved in the hands of scholars attentive to the varied ways communities are creating datasets within their own terms. Considering this, the concept can also assist in how we rhetorically talk about data structures that exist beyond strict quantitative approaches. We've learned that the flattening of people, often victims of hegemonic violence, begets statistics that are often incorrect or even nonexistent. Corrective actions against this often implore the lived experiences and rich narratives of communities. Counterdata positions memory, place, and storytelling as viable resistance strategies against state control and legibility. Even though counterdata cannot be the sole means by which the subjugated emancipate themselves, counterdata production brings visibility to important methodologies that can hopefully continue promoting the narratives communities create for themselves.

22 Annita Hetoevehotohke'e Lucchesi, "Mapping Violence against Indigenous Women and Girls: Beyond Colonizing Data and Mapping Practices," *ACME: An International Journal of Critical Geographies* 21, no. 4 (May 2022): 389–98; Stephanie Russo Carroll, Marisa Duarte, and Max Liboiron, "Indigenous Data Sovereignty," in *Keywords of the Datafied State,* eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

COUNTERDATA
COUNTERDATA
COUNTERDATA

BY
VANESSA MASSARO, DARAKHSHAN J. MIR,
TERRELL MOSLEY, and NATHAN C. RYAN

COUNTERDATA
COUNTERDATA
COUNTERDATA

COUNTERDATA

By Vanessa Massaro, Darakhshan J. Mir,Terrell Mosley, and Nathan C. Ryan

Clients who find themselves in need of Terrell's services regularly confront a supervision system that ensnares them more than it sets them up for even the system's definition of success.

Dawn was a working professional, who, enveloped in an abusive romantic relationship, developed a substance use disorder. When she was initially arrested for drug possession, she was sentenced to probation. She was not able to stop using drugs and eventually violated her probation with a "hot urine." This resulted in jail time and extended probation. Now Dawn is home again with two kids to support, probation time to serve, and she, due to her incarceration, no longer has her professional license.

Jessi's[1] life seemed to veer off track in his teenage years shortly after his parents' divorce. His mother suffered a substance use disorder and Jessi started using drugs at an early age. He was first charged as a teenager with possession of drug paraphernalia. Jessi completed his assigned Accelerated Rehabilitative Disposition[2] program successfully, but violated probation with a positive drug test and was sent to jail. He is 20 now and has never been free from supervision, cycling through jail, prison, rehab, and supervision. He suffers from mental health and substance use disorders that have not shown much improvement. He needs a wide range of support and services that are not readily available.

1 Jessi is a composite of three clients with similar trajectories to maintain confidentiality.

2 Accelerated Rehabilitative Disposition is an option for first time offenders to complete a rehabilitative program and have their charges dismissed and records expunged.

Introduction

In what follows, we use recidivism as an entry point to form an empirically grounded conceptualization of counterdata projects that challenge the datafied carceral state. We follow Seyi Olojo's definition of counterdata, which she defines as data "that is collected in contestation of a dominant institution or ideology," and "is collected as a means for communities to tell their own stories through the use of data."[3] We choose to interrogate recidivism because it is datafied by the Pennsylvania Department of Corrections (PA DOC), it embodies the DOC's own metric of assessing success,[4] and it is a concept that the datafied state itself marks as simultaneously insufficient *and* determining of people's fates, thereby providing a rich case study to conceptualize a counterdata project.

In the context of the US criminal punishment system, we advocate for counterdata and counteranalyses that hold state entities accountable and center the needs and aims of incarcerated people. In that regard, counterdata should highlight the fallacies of the system, center the victims of state violence, and/or reduce harm.[5] Counteranalyses are analyses that either produce counterdata (even if they take state data as input) or are analyses that take counterdata as input. We pay attention to both counterdata and counteranalysis as two intertwined aspects of challenging the datafied carceral state. We distinguish between counterdata and counteranalysis to capture the countering potentials of both the data as well as the analyses. It may be the case that while underlying data are produced by the state and its associated entities, the countering could occur by outside agents, organizations, and/or activists seeking to break the hegemonic grasp of the state on such data through the production of counterdata and/or counteranalyses.

3 Seyi Olojo, "Counterdata," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

4 Kristofer Bret Bucklen, Michele Sheets, Chloe Bohm, Nicolette Bell, Jessica Campbell, Robert Flaherty, and Kate Vander, "Recidivism 2022 Report," Pennsylvania Department of Corrections, November 2022, https://www.cor.pa.gov/About%20Us/Statistics/Documents/Reports/Recidivism%202022%20Report.pdf.

5 Catherine D'Ignazio and Lauren F. Klein, Data Feminism (Cambridge: MIT Press, 2020); Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016); Julia Angwin, Jeff Larson, Surya Matttu, and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=DtgTx_YLhwojQCM_xkrr4my1nl7Ucetj; Sandra Harding, *Whose Science? Whose Knowledge?: Thinking from Women's Lives* (New York: Cornell University Press, 1991).

The opening vignettes capture Terrell Mosley's experiences in his work as a reentry coordinator for Susquehanna Valley Mediation in Selinsgrove, PA. His position exists as a direct response to needs identified by formerly incarcerated people in the Susquehanna Valley of central Pennsylvania. When we examine the wide range of Mosley's clients, patterns of trauma, poor mental health, and isolation from the mainstream economy emerge.[6] Clients find themselves without the skills, resources, and tools to get their lives onto the track diversion and supervision demand. When they find themselves unable to do so they are promptly returned to prison in a neoliberal manifestation of the criminal punishment system's underlying logic of personal responsibility.[7] The dominant assumption in carceral systems is that one's *individual* conditions are predictors of success, yet these stories reveal predictable patterns of *institutional* failure, and the institution's inability to set people up for its own definition of success. If Jessi tests positive for drugs, did Jessi or his probation officer fail? Or was the Accelerated Rehabilitative Disposition program a failure?

In his position, Mosley, one of the authors of this piece, draws from his personal experience as a Black man growing up in central Pennsylvania being targeted by police and eventually incarcerated through a plea deal and probation violation. The countering he engages in emerges from his community-based work and experience and informs the academic study of the other three authors, Vanessa Massaro, Darakhshan Mir, and Nathan Ryan, who comprise an interdisciplinary research team that studies carceral data and algorithms through a critical data studies lens.

In 2018, the team requested data from the PA DOC that contains variables related to parole decisions for more than 280,000 distinct individuals. The simultaneously expansive, intrusive, and reductive dataset consists of over

6 Susan Dewey, Bonnie Zare, Catherine Connolly, Rhett Epler, and Rosemary Bratton, *Outlaw Women: Prison, Rural Violence, and Poverty in the New American West* (New York: NYU Press, 2019); Kimberle Crenshaw, "From Private Violence to Mass Incarceration: Thinking Intersectionally About Women, Race, and Social Control," *UCLA Law Review* 59, no. 1418 (September 1, 2012): 1420–72; Beth E. Richie, *Arrested Justice: Black Women, Violence, and America's Prison Nation* (New York: NYU Press, 2012).

7 Judah Schept, *Progressive Punishment: Job Loss, Jail Growth, and the Neoliberal Logic of Carceral Expansion* (New York: NYU Press, 2015).

1,200 variables about the lives and experiences of incarcerated people. This includes parole board decisions, custody levels, demographic information, disciplinary actions while incarcerated, mental health diagnoses, and their physical movement through the PA DOC. In prior work, team members have used this data to counter the carceral state's narrative about the rehabilitative capacities of their tools and processes.[8] In this work, Massaro, Mir, and Ryan combine their previous experience with carceral data and Mosley's experience to drive a conception of counterdata and counteranalyses (see Table 1).

What Does Data Counter and How?

Drawing from Mosley's experience, we note that there is widespread recognition among incarcerated people that probation and correctional officers possess a great deal of discretion when making decisions about parole and disciplinary violations, respectively. A collection of community-based data from the clients Mosley works with regarding the perceived fairness of such officers is crucially missing data[9] — it is also an example of counterdata. Subsequently analyzing these data for a correlation with the parole outcomes of individuals would constitute a valuable counteranalysis, shifting the variable of inspection from the incarcerated individual to the officer. Another example of counteranalysis is evaluating the population density of an incarcerated person's state correctional institute (indicating their experience of overcrowding) and how likely density shapes their mental health and behavior. The PA DOC collects this data, but does not consider this data when predicting the likelihood of the incarcerated person recidivating through what they call a risk survey instrument. Their oversight eschews any concerted consideration of mental health status into actual outcomes

8 Vanessa A. Massaro, "Relocating the 'Inmate': Tracing the Geographies of Social Reproduction in Correctional Supervision," *Environment and Planning C: Politics and Space* 38, no. 7–8 (November 2020): 1216–36, https://doi.org/10.1177/2399654419845911; Andrew G. Ferguson, "Policing Predictive Policing," *Washington University Law Review* 94, no. 5 (January 2017), https://journals.library.wustl.edu/lawreview/article/id/3851/; Robert R. Belair, Paul L. Woodard, and Eric C. Johnson, "Use and Management of Criminal History Record Information: A Comprehensive Report," US Department of Justice, 2001, https://bjs.ojp.gov/content/pub/pdf/umchri01.pdf; Wayne Logan and Andrew Ferguson, "Policing Criminal Justice Data," *Minnesota Law Review* 101, no. 2 (December 2016): 541–616.

9 As an entry point to counterdata, we can consider what type of data is not collected by the state and what kinds of analyses are not undertaken by the state as part of its decision-making apparatus. See: Sharlene Nagy Hesse-Biber, *Handbook of Feminist Research: Theory and Praxis* (Thousand Oaks: SAGE, 2012); Brittany Farr, "Witnessing an Absent Presence: Bringing Black Feminist Theory to Traditional Legal Archives." *Black Scholar* 52, no. 4 (2022): 64–75, https://doi.org/10.1080/00064246.2022.2111676. See Alessandra Jungs de Almeida, Lauren Klein, and Catherine D'Ignazio, "Missing Data" in Keywords of the Datafied State, eds. Ranjit Singh, Jenna Burrell, and Patrick Davison. That entry also credits the artist Mimi Ọnụọha.

for people while simultaneously signaling the importance of mental health status to parole and recidivism outcomes.[10] We argue that counterdata and counteranalyses should produce a vision from outside the carceral state, that should center the experiences of (formerly) incarcerated people and shift the responsibility to institutions.

Counterdata Analysis of Recidivism

Recidivism is a long-standing binary variable in the criminal punishment system. It is typically measured at one year and three years post-release. Recidivism rates are calculated based on the percentage of people who return to prison within that respective time. The rate is meant to evaluate the success of people post-release. Despite nearly a century of collecting incredibly detailed, to the point of invasive, data, the recidivism rate remains stagnant.[11] What, precisely, does the recidivism rate assess — individuals or the correctional system?

Recidivism exemplifies a strategy for enacting counterdata because it could be framed as a measure to evaluate the corrections system. The carceral state's focus on reducing recidivism rates indicates their awareness of the correctional system's failure to rehabilitate people. However, the state is invested in attempts to predict the portended risk that an *individual* has of recidivating through the use of the Sentencing Risk Assessment Instrument,[12] thereby placing the onus on the individual. There is little effort to assess and change institutional practices to reduce structural causes of recidivism, let alone examine the role the system itself plays in individuals recidivating. The missing counteranalysis reframes recidivism to be a failure of the system and its policies rather than the individuals passing through it.

10 Pennsylvania Department of Corrections, "Recidivism Report 2022," https://www.cor.pa.gov/About%20Us/Statistics/Documents/Reports/Recidivism%202022%20Report.pdf

11 Pennsylvania Department of Corrections, "Department of Corrections Procedures Manual: Reception and Classification,"; Pennsylvania Board of Probation and Parole, "Annual Report 2016," https://www.parole.pa.gov/Information/publications/Documents/PBPP%2016%20AR%20FINAL.pdf.

12 This predictive analysis undertaken by the state is based on variables collected by the state (such as age, gender, number of prior convictions, prior conviction offense type), and owes its existence to a long history of extreme datafication of the carceral experience in the United States, including the PA DOC. See: Brian Jefferson, *Digitize and Punish: Racial Criminalization in the Digital Age* (Minneapolis: University of Minnesota Press, 2020).

From Recidivism to Desistance

Recently, state actors have moved away from recidivism as an assessment variable due to its narrow scope.[13] The National Institute of Justice (NIJ) instead recommends tracking "desistance," defined as "an individual's progress toward moving away from crime."[14] This emerging guidance corroborates years of scholarship that identifies, from myriad angles, the significant and impactful role of institutional decision-making on individual outcomes[15]. Pennsylvania has followed suit: in the PA DOC's 2022 Recidivism Report, the agency pushes for the use of desistance to better show the institution's "success."

Even with this move to a different assessment variable, the state's evaluations consistently place the failures on the incarcerated person rather than examining recidivism communally or systemically. This is because desistance as promoted by the NIJ remains highly focused on individuals being "successful." Desistance remains anchored on the concept of criminality (the propensity of an individual to offend) and thus continues to assess the individual rather than institutions. Desistance is therefore an attempt at a progressive reform of recidivism data but is not productive of counterdata or a counteranalysis. For that, a greater shift in the focus of measurement is required — one that turns accountability on the institutions, not individuals.

The state's failure to consider the institutional, contextual, or systemic factors that impact a person's likelihood to recidivate produces a wealth of missing data. For example, the PA DOC does not collect data on probation and parole officers' disposition, background, mental health, drug use, or any of the other myriad variables impacting successful performance of their jobs. Yet, all these variables are collected for incarcerated people during and after incarceration. We are not suggesting that such a collection by the state would remedy the underlying problems, but rather we are drawing attention to the contours of a (counter)-datafication project of the institution and institutional actors.

Another example is data collected on transitional housing. In Pennsylvania, these are either halfway houses or state contracted community corrections centers (CCCs). There is no data on ownership structures, programming, or financial policies. Homing in on financial policies of CCCs,

13 Nathan James, *Offender Reentry: Correctional Statistics, Reintegration into the Community, and Recidivism* (Washington: Congressional Research Service, 2015), https://sgp.fas.org/crs/misc/RL34287.pdf; Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," Science Advances 4, no. 1 (2018), https://doi.org/10.1126/sciadv.aao5580.

14 National Institute of Justice, *Desistance from Crime: Implications for Research, Policy and Practice* (Washington: National Institute of Justice, 2021), https://www.ojp.gov/pdffiles1/nij/301497.pdf.

15 Angèle Christin, Alex Rosenblat, and danah boyd, "Courts and Predictive Algorithms" (workshop, Data & Civil Rights: A New Era of Policing and Justice, Washington, DC, October 13, 2015), https://www.datacivilrights.org/pubs/2015-1027/WDN-Predictive_Policing.pdf; Mikaela Meyer, Aaron Horowitz, Erica Marshall, and Kristian Lum, "Flipping the Script on Criminal Justice Risk Assessment: An Actuarial Model for Assessing the Risk the Federal Sentencing System Poses to Defendants," *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, (New York, June 2022): 366–78. https://doi.org/10.1145/3531146.3533104; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018); Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford: Polity, 2019); Jackie Wang, *Carceral Capitalism* (South Pasadena: Semiotext, 2018).

there is no consistent data on how money is collected and held by a specific house and the fees charged to residents. In Massaro's ethnographic work in Philadelphia, she often heard stories of residents' money and paychecks going unaccounted for in these institutions. How might these data be better collected to hold those institutions accountable to formerly incarcerated residents, and thereby society at large? Further, how might being a victim of such an incident impact your inability to get your life on track and your likelihood to recidivate? Such counterdata and counteranalyses guide different questions toward different systemic ends.

Further, not all data is quantitative. What would a qualitative and holistic version of recidivism look like? Could it be understood and evaluated through storytelling? How long does it take you to get back on your feet? What do you need to get back on your feet? How much time do you need to get your feet under you? These are essential qualitative questions that also counter recidivism as a binary variable and represent a move toward "thick data."[16]

We can also examine the wide range of already existing data that can be counteranalyzed to serve different ends. Recidivism is a binary calculation of individuals. While more recent policy literature marks a turn toward desistance from crime, the main variables are still calculated on individuals: deceleration (slowing the rate of offending), de-escalation (reducing the seriousness of offenses), and cessation (the stopping of offending altogether).[17] Using these variables, the PA DOC found that nine out of ten DOC reentrants meet one or more of these measures — the report does not ask why, if this is the case, 50 percent of them still return to prison. Desistance fails to be an

16 Tricia Wang, "Why Big Data Needs Thick Data," Medium, December 5, 2016, https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7.

17 Bucklen et al., "Recidivism Report 2022."

example of counterdata as it is being collected by the datafied state itself and it is decidedly not an example of counteranalysis since desistance only has meaning when compared to recidivism and therefore demands that our attention be placed on recidivism.

A counterdata approach can reorient variables already collected by the state to serve different ends, namely, support, healing, and rehabilitation. This radical reorientation of the goals and purposes of these analyses produces the resulting counterdata and counteranalyses. Table 1 lists several examples.

Table 1. Examples of Counterdata and Counteranalyses

The PA DOC …	Community organizations/individuals while enacting counterdata projects could …
Collects data about positive drug tests	Analyze how accessible substance abuse health care is and use that to contextualize a positive drug test in order to advocate for an incarcerated person or someone on parole.*
Reports recidivism rates	Use it as a measurement of institutional failure instead of as a measure of personal failure.
Monitors employment status	Analyze how easy it is to secure employment for someone in the parolee's neighborhood and with their history with the PA DOC to contextualize their employment status and advocate for them.*
Incarcerates parole violators	Assess the effect a parole officer has on a parolee violating the terms of their parole.
Prohibits contact between a parolee and the formerly incarcerated	Understand the family and social networks of the parolee to determine the benefit of having such networks, even if they consist of formerly incarcerated people.
Collects detailed data on PA DOC visitors, including drug searches	Determine the social and financial pressures incarceration has for a person's support network and offer increased services to the people who help incarcerated people during and after their period of incarceration.
Collects data on mental health to determine custody (housing security) levels	Assist incarcerated people with mental health care such as help in getting a dual diagnosis (both mental health and substance abuse disorder) for a better housing outcome.* Collect data to analyze the relationship between dual diagnoses and housing outcomes.
Collects data on race of incarcerated people	Analyze how the minoritized status of incarcerated people in a correctional institute (relative to its staff) impacts their experiences of incarceration — such as disciplinary tickets, indicating friction with the correctional staff — and ultimately parole outcomes.
Collects data on the capacity and population of each correctional institute	Analyze how the extent of overcrowding impacts the mental health and behavior of people. Analyze political and economic incentives to continue to populate prisons and stuff beds. Create an institutional score for each correctional institution that reflects their failure at rehabilitating people.
Does not collect data on the behavior and professionalism of parole or probation officers	Collect data on the behavior and professionalism of parole and probation officers. Analyze this data for the impact on people's parole outcomes.
Does not publicly report data on the success of cases assigned to a probation officer	Collect and publicize data (via public record requests and/or community-based collection) on the outcomes of cases assigned to individual probation officers.

* Work that Mosley engages in as reentry coordinator at Susquehanna Valley Mediation.

* Acknowledgment: This work was partially supported by National Science Foundation Award #2213826.

Conclusion

The carceral state's processes of data collection are dehumanizing. Recidivism reduces people to binaries and even the most supportive, restorative parts of the incarcerated experience: such as visitations, your number of visits, your mail (which is read and heavily surveilled), your phone calls (also datafied and heavily surveilled), and your mental health notes; all become part of the database — further reducing people to numbers. Mosley's clients help us consider the possibilities of developing a counteranalysis of the variable and targets for evaluation.

We propose counterdata as a more comprehensive way of challenging the status quo, by not only patching gaps and omissions in the data but also challenging the datafied state's analyses and/or data. We imagine a system that could be focused on counteraction that leads us to healing and justice. While institutional data and algorithms seek to entrench punitive priorities of the state (under the veneer of neutrality), counterdata seeks to challenge the state's power and move the world toward what it could be, what it should be. When these larger visions of healing, restoration, and justice are centered, we can consider the larger goals of counterdata and missing data: that of examining institutions for their role in manufacturing harm, enabling us to flip the script of evaluation. What data would be needed to calculate a bank score (capturing how risky a bank is for an individual) rather than a credit score (how risky an individual is for a financial institution)? What would an institutional score for each prison in the Department of Corrections create instead of a score for each incarcerated person? Counterdata and connected counteranalyses could work in service of institutional accountability, transparency, and ultimately a larger reimagination in service of restoration and liberation.

MISSING
MISSING
MISSING

BY
ALESSANDRA JUNGS DE ALMEIDA, LAUREN KLEIN,
and CATHERINE D'IGNAZIO

DATA
DATA
DATA

MISSING DATA

By Alessandra Jungs de Almeida, Lauren Klein, and Catherine D'Ignazio

Introduction

There are several different definitions of missing data. While some might refer to data that is literally absent, as in statistical approaches to missing data that attempts to interpolate what might fill in the gaps,[1] others, such as the artist and educator Mimi Ọnụọha, take "missing data" to mean something more political — "something [that] does not exist, but it should."[2] In the same line as Ọnụọha, our definition of missing data refers to information that goes uncounted (or otherwise unrecorded), despite social and political demands that such data should be collected and made available. Our concept of missing data may include entirely absent data, as well as data that is sparse, neglected, poorly collected and maintained, purposely removed, difficult to access, infrequently updated, contested, and/or underreported.[3]

Missing data, in the expanded definition we propose in this essay, is a political concept. On one hand, missing data can function as a challenge from civil society to formal institutions, including governments, religious institutions, and corporations. In these cases, it represents a demand from specific communities about public issues that concern society writ large. On the other hand, missing data may be actively desired and produced by marginalized groups seeking to protect information about their community and culture from the eyes of institutions. In these cases, the data is "missing" for institutions, which make a demand for information that is actively

1 Dirk Temme and Sarah Jensen, "Missing Data – Better 'Not to Have Them', but What If You Do? (Part 1)," *Marketing: ZFP – Journal of Research and Management* 41, no. 4 (2019): 21–32, https://www.jstor.org/stable/26873605.

2 Mimi Ọnụọha, "On Missing Data Sets," GitHub, January 24, 2018, https://github.com/MimiOnuoha/missing-datasets; Mimi Ọnụọha, "The Library of Missing Datasets," Mimionuoha.com, https://mimionuoha.com/the-library-of-missing-datasets.

3 Catherine D'Ignazio, *Counting Feminicide: Data Feminism in Action* (Cambridge: MIT Press, 2024); Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge: MIT Press, 2020).

protected by and kept within a community. In this sense, missing data is also a relational concept because it implies a directionality — an informatic demand *from* one group or institution *to* another group or institution. Missing data is not always a bad thing, nor always a good thing. Instead of thinking of it normatively, the locus of analysis should be on the social context, who is making the demand to whom, and the political context for which specific information is deemed to be missing. Our definition differs from other more technical notions of missing data that may not consider or highlight the unbalanced power relationships between different social actors, such as marginalized communities and the state. In this sense, the definition of missing data proposed here explicitly includes a political demand, because the group making the demand for information is trying to charge another group or institution with the responsibility for the absence of this data. When this relates to marginalized groups making demands on the state, groups are also trying to assert the institutional neglect of the group or issue represented by the data. Given the focus on the datafied state, this article will focus particularly on missing data related to governments, where civil society groups demand that the government collect specific data or where the government demands data that communities seek to protect.

Missing Data and the State

Different examples worldwide illustrate missing data's relational and political characteristics in the state-society interaction. For instance, during the second half of the 20th century, social movements in Latin America struggled to find data on the disappeared people whom authoritarian governments had arrested and often tortured or killed. One of the more significant examples of these struggles happened in Argentina. Since the 1970s, civil

society organizations have pushed for comprehensive records of the people detained by the Argentine government. This action came especially from the organizations *Comitê de Defesa dos Direitos Humanos para os países do Cone Sul* (CLAMOR), *Madres de la Plaza de Mayo*, and *Familiares de Desaparecidos y Detenidos por Razones Políticas.* In a 1980 public petition, these groups demanded that the military government "publish the list of the detained-disappeared, where they are and the reason for their detention."[4] Although the country was still under a dictatorship, the document had more than 12,000 signatures, and members of the media counted about 500 people protesting on the day the organizations delivered the document to the military government.[5] This collective action and the political demand for state answers eventually resulted in a state list with 8,961 names of disappeared people.[6] The Argentine state conducted this search during its democratic transition (1983 and 1984) and organized its search based on a list of 7,000 names that the organization CLAMOR, based in Brazil, had previously collected.[7] The final list published by the government, which is utilized to this day, is still considered incomplete and contested by civil society. Organizations like *Madres y Abuelas de la Plaza de Mayo* advocate that there were 30,000 disappeared, and the organization Comunidad Homosexual Argentina (CHA) claims 400[8] LGBTQIA+ people also disappeared and were not in the CONADEP report.[9]

At the same time, with the 8,961 names, the democratic Argentine state, with the strong influence of these movements, was able to work with civil society to start a process for justice for the disappeared people and their families. In this sense, initially, the identification and collection of this missing data began as a collective process for justice in the face of state authoritarianism. This was undertaken with the help and collective memory of the entire Argentine society demanding data on the disappeared people by governmental institutions.[10] If not for the social movements' demand for

4 Ulises Gorini, *La rebelión de las Madres: Historia de las Madres de Plaza de Mayo Tomo I* (1976–1983), (Buenos Aires: EDULPD, 2017).

5 Gorini, *La rebelión de las Madres*.

6 "Part VI. Recommendations and Conclusions," CONADEP, September 1984, http://www.desaparecidos.org/nuncamas/web/english/library/nevagain/nevagain_283.htm

7 Ana Célia Navarro de Andrade and Heloísa de Faria Cruz, *Clamor e Ditaduras no Cone Sul: Documentacao, Memoria e Pesquisa* (Brazil: EDUC, 2021); CLAMOR, "O boletim Clamor," PUC-SP, Accessed March 2, 2023, from http://www.pucsp.br/cedic/clamor.

8 These numbers resulted in a new public claim from different civil society organizations in Argentina: "¡30.400 *presentes!*" This claim is related to the recognized claim from *Madres y Abuelas,* about 30.000 disappeared people.

9 Abuelas de Plaza de Mayo, "Historia | Las Abuelas | Abuelas de Plaza de Mayo," retrieved April 19, 2023, from https://www.abuelas.org.ar/abuelas/historia-9; Redacción Clarín, "Una Duda Historica: No se Sabe Cuantos Son Los Desaparecideos," *Clarín*, May 10, 2003, https://www.clarin.com/ediciones-anteriores/duda-historica-sabe-desaparecidos_0_B1FG1JglCKl.html.; Emmanuel Theumer, Noelia Trujillo, and Marina Quintero, "El Nunca Mas de los 400: Políticas de Articulación del duelo y la Reparación en la Argentina Reciente," *El lugar sin límites*, no. 3 (April 2020): 48–64.

the missing data, information on the disappeared people would never have been available. Consequently, processes of memory for their lives and justice for their families would never have been made.

Cases such as Argentina's are not unique. Other missing data cases related to citizens' political demands on the state include topics such as violence against women,[11] accessibility in the urban environment,[12] data on children with microcephaly resulting from the Zika virus,[13] data that tracks gun violence in the US,[14] as well as projects that point out gaps in environmental data, evictions data tracking, police killings of citizens, and maternal mortality cases, among many others. In all cases, the absence of certain data points indicates a lack of political determination in collecting this data, which results in civil society groups questioning the states' inaction and demanding accountability through data collection and distribution.

Missing Data as Strategic Neglect from the State and Powerful Institutions

Because producing data requires resources — to acquire, maintain, publish and use — missing data can result from resource allocation decisions made by powerful institutions, such as companies, intergovernmental organizations, religious institutions, and especially governments. Therefore, demanding missing data is a civil society strategy to hold the datafied state accountable for what is missing and, crucially, why. In this context, calling attention to missing data, as well as other absences, can be a strategy for unsettling whoever is trying to forget or hide information that could demonstrate the inequalities of our societies.[15]

10 Laura Marina Panizo, "Muerte, Desaparición y Memoria: El Caso de Los Desaparecidos de La Última Dictadura Militar En Argentina," *Historia, Antropología y Fuentes Orales*, no. 42 (2009): 71–84. http://www.jstor.org/stable/25759001.

11 Helena Suárez Val, Sonia Madrigal, Ivonne Ramírez Ramírez, and María Salguero, *"Monitoring, Recording, and Mapping Feminicide — Experiences from Mexico and Uruguay," in Femicide Volume XII Living Victims of Femicide,* eds. Helen Hemblade and Helena Gabriel (Vienna: UNSA, 2019), 67–73.

12 Shiloh Deitz, Amy Lobben, and Arielle Alferez, "Squeaky Wheels: Missing Data, Disability, and Power in the Smart City," *Big Data & Society* 8, no. 2 (July 2020): 20539517211047735, https://doi.org/10.1177/20539517211047735.

13 Debora Diniz, *Zika: Do Sertão nordestino à ameaça global: Do Sertão Nordestino à Ameaça Global* (Brazil: Civilização Brasileira, 2016).

14 Gun Violence Archive. Evidence Based Research — since 2013. *https://www.gunviolencearchive.org/*

15 Debora Dinz, "Lembrar" *in Esperança Feminista (2a edição),* eds. Debora Diniz and Ivone Gebara (Brazil: Rosa dos Tempos, 2022).

As a case in point, we might consider the data regarding people annually killed by the police in Brazil. Each of the 26 Brazilian states collects this data, but they do not make it publicly available. Through a law on information access, an initiative conducted by the media and an organization called *Fórum de Segurança Pública* regularly requests data from the Brazilian states.[16] This initiative then publishes the data. Thanks to the group, it is now possible to compare the number of killings with other countries' data and find evidence at scale for how the Brazilian police force is one of the most lethal in the world.[17]

Yet the clarity offered by such a dataset on police killings in Brazil still cannot account for the full extent of state violence. The Brazilian police also participate in racist violence, among other structural oppressions. In 2021, 11 of the 26 Brazilian states did not collect (or decided not to share) data on the race of the people killed by the police.[18] Based on the existing data from other states, it is possible to determine that 81.5 percent of the victims were Black. While data alone will not solve racialized police violence,[19] the political decision to not collect or not publish this data reflects a desire on the part of the state to maintain a violent order by neglecting "data and statistics about those minoritized bodies who do not hold power."[20] In this sense, neglecting to collect this data is strategic because it allows the state to maintain a public security policy in which killing and torturing Black people is possible without having public scrutiny on the actual scope and scale of these killings.[21]

Beyond the negligence in collecting data, there are illustrative examples of active political choices by governments to produce missing data. This is to say that data that was previously published is unpublished, taken down, disappeared, removed or deleted. For instance, since the COVID-19 pandemic started, the disappearance of data on the number of people with the

16 Clara Velasco, Felipe Grandin, and Alessandro Feitosa Jr., "Número de Pessosas Mortas Pela Polícia Cai e Atinge Menor Patamar Em Quatro Anos; Assassinatos de Policiais Também Têm Queda," *G1*, May 4, 2022, https://g1.globo.com/monitor-da-violencia/noticia/2022/05/04/numero-de-pessoas-mortas-pela-policia-cai-e-atinge-menor-patamar-em-quatro-anos-assassinatos-de-policiais-tambem-tem-queda.html

17 Samiro Bueno and Beatriz Rodrigues, "Letalidade Policial: Uma Resiliente Prática Institucional," in *Fórum Brasileiro de Segurança Pública,* https://forumseguranca.org.br/storage/8_anuario_2014_20150309.pdf

18 Clara Velasco, Alessandro Feitosa Jr., and Felipe Grandin, "11 Estados Não Divulgam Dados Completos de Raça de Mortos Pela Polícia; Números Disponíveis Mostram Que Mais de 80% Das Vítimas São Negras," G1, May 4, 2022, https://g1.globo.com/monitor-da-violencia/noticia/2022/05/04/11-estados-nao-divulgam-dados-completos-de-raca-de-mortos-pela-policia-numeros-disponiveis-mostram-que-mais-de-80percent-das-vitimas-sao-negras.ghtmln.

19 Catherine D'Ignazio and Lauren F. Klein, "Seven Intersectional Feminist Principles for Equitable and Actionable COVID-19 Data," *Big Data & Society* 7, no. 2 (July 2020): 2053951720942544, https://doi.org/10.1177/2053951720942544.

20 D'Ignazio and Klein, *Data Feminism.*

21 Geledés, 'Existe, Por Parte do Estado Brasileiro, Uma Política de Extermínio da População Negra," *Geledés*, April 19, 2023, https://www.geledes.org.br/existe-por-parte-do-estado-brasileiro-uma-politica-de-exterminio-da-populacao-negra/.

virus and the deaths related to it has been denounced in many countries, including Brazil and the US.[22] Furthermore, data from the US Environmental Protection Agency (EPA) during the Trump administration also actively disappeared from its website.[23] The Trump administration deleted pages on climate change, downplayed references to it in different federal reports and policies, and removed access to 20 percent of its website.[24] The fear among scientists and activists that this data disappearance would escalate was the motivation for some archival initiatives to preserve the environmental data from EPA in the US. For instance, the Environmental Data & Governance Initiative archived 200 terabytes of data from government websites between 2016 and 2017.[25] Scientists and advocates asserted that these actions deterred the administration from deleting all federal environmental data, an example of how archival practices can perform activist work.[26]

These examples demonstrate how political demands for missing data, as we propose here, can operate as a theory of change that more information can shed light on social inequalities and injustices, affecting the daily lives of populations. They also exemplify the political content of the missing data concept, such as its relational nature — how interactions and demands between specific groups and powerful institutions are an essential part of the concept, encompassing both data requests and protection from civil society groups. Typically, it is civil society — including NGOs, journalists, activists, scientists, academics, and community-based organizations — that pressures governments to make sure measures are taken to count and publish data responsibly and fairly, often as a method to address racial, gender, and class inequalities, such as in the cases handled above. In themselves, these actions are embedded in the understanding of the unbalanced power relations in our societies and seek to hold the datafied state accountable.

22 Marina Novaes, "Governo Bolsonaro Impõe Apagão de Dados Sobre a COVID-19 no Brasil em Meio à Disparada das Mortes," *El País Brasil,* June 6, 2020, https://brasil.elpais.com/brasil/2020-06-06/governo-bolsonaro-impoe-apagao-de-dados-sobre-a-covid-19-no-brasil-em-meio-a-disparada-das-mortes.html.; Lena H. Sun and Amy Goldstein, "Disappearance of COVID-19 Data from CDC website Spurs Outcry," *Washington Post*, July 16, 2020, https://www.washingtonpost.com/health/2020/07/16/coronavirus-hospitalization-data-outcry/.

23 Eric Nost, Gretchen Gehrke, Grace Poudrier, Aaron Lemelin, Marcy Beck, and Sara Wylie, "Visualizing Changes to US Federal Environmental Agency Websites, 2016–2020," *PLOS ONE* 16, no. 2 (February 2021): e0246450, https://doi.org/10.1371/journal.pone.0246450.

24 Dino Grandoni and Brady Dennis, "Biden Administration Revives EPA Web page on Climate Change deleted by Trump," *The Washington Post, March 18*, 2021, https://www.washingtonpost.com/climate-environment/2021/03/18/epa-website-climate/; Nost et al., "Visualizing Changes."

25 Justine Calma, "How Scientists Scrambled to Stop Donald Trump's EPA from Wiping Out Climate Data," *The Verge*, March 8, 2021, https://www.theverge.com/22313763/scientists-climate-change-data-rescue-donald-trump.; "Our Story," Dta Refuge, retrieved April 21, 2023, from https://www.datarefugestories.org/our-story-1.

26 Calma, "How Scientists Scrambled."

Missing Data and Counterdata Production

Once identified as such, missing data can become a focal point for political action, including but not limited to the demand for data production and availability. This is well illustrated by the examples of the US EPA and Brazil's police force. In both cases, research institutes, activists, scientists and/or the media worked to produce data, protect it, and make it available. When these actions are not enough, the identification of missing data can mark the beginning of a longer political process whereby members of civil society take action to produce missing data themselves. This action is often described in terms of counterdata collection. In *Keywords of the Datafied State*, Seyi Olojo defines counterdata as "data that is collected in contestation of a dominant institution or ideology," also being "a means for communities to tell their own stories through the use of data."[27] We echo that definition here, emphasizing how the counterdata collection process is active, intentional, and contextualized, undertaken with political aims and usually in explicit relation to the state or other powerful institutions, which will be called upon to address those public aims.

We can understand what happened in the Argentine case, when CLAMOR gathered data on the disappeared people in order to challenge the state's absence of data, as an example of counterdata production. Another demonstration of counterdata production is found in family-led organizations in Mexico that collect data on the disappeared, primarily led by mothers searching for their children.[28] The Mexican government lacks comprehensive data on disappearances and does not prioritize searching for the disappeared, despite at least 100,000 open cases.[29] In a counterdata action, these organizations conduct independent research, create lists of the

27 Seyi Olojo, "Counterdata" in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society Research Institute, 2024).

28 Arely Cruz-Santiago, "Lists, Maps, and Bones: The Untold Journeys of Citizen-Led Forensics in Mexico," *Victims & Offenders* 15, no. 3 (April 2020): 350–69, https://doi.org/10.1080/15564886.2020.1718046.

29 Vanessa Buschschlüter, "Mexico Disappearances Reach Record High of 100,000 Amid Impunity," *BBC News*, May 17, 2022, https://www.bbc.com/news/world-latin-america-61477704

disappeared, search for clandestine mass graves, and preserve bones to be further analyzed, what Arely Cruz-Santiago defines as a scientific act of care and a political act at the same time.[30]

Counterdata production can also describe what happens in the Americas (and other places around the world) when feminist data activists collect data on the killings of women — including cis and trans women and other feminized bodies — because their governments undercount or do not count these killings at all.[31] To illustrate the strength of these counterdata activists, the Data + Feminism Lab at MIT has mapped more than 180 organizations worldwide that produce counterdata on lethal violence related to gender. Many of these organizations produce their data to contest the state's missing data about gender-based violence. This counterdata (and related actions) pressures the state for additional data availability, financial resources, and/or public policies and political interventions, refusing to allow the government's continued negligence of these issues.

A 2020 report from *Socorristas en Red (feministas que abortamos),* a Argentine organization, exemplifies the impact that counterdata production can have on the state, including laws and policies. Until 2020, voluntary abortion was not legalized in the country, thus there was scarce official data on the reality of abortions. Between 2014 and 2019, this organization supported and interviewed 38,116 women in Argentina, offering them information on the use of medication for abortion.[32] In the report, the organization shared the results of these interviews, presenting an aggregated look at the experiences of women and trans people they help to perform abortions.[33] Through their counterdata production, *Socorristas en Red* produced robust evidence on the reality of abortion in the country, supporting arguments for the ongoing political debate on the legalization of abortion. Together with

30 Cruz-Santiago, "List, Maps, and Bones."

31 D'Ignazio, *Counting Feminicide*; Catherine D'Ignazio, Isadora Cruxên, Helena Suárez Val, Angeles Martine Cuba, Mariel García-Montes, and Silvana Fumega, Harini Suresh, and Wonyoung So, "Feminicide and Counterdata Production: Activist Efforts to Monitor and Challenge Gender-Related Violence," *Patterns* 3, no. 7 (July 2022), https://doi.org/10.1016/j.patter.2022.100530.; Suárez et al., *Monitoring, Recording, and Mapping Feminicide*.

32 Socorristas en Red. "Sistematización de Acompañamientos a Abortar: Realizados En El Año 2019 Por Socorristas En Red (Feministas Que Abortamos)." Argentina: Socorristas en Red, April 2020. https://socorristasenred.org/sistematizacion-2019/.

33 Socorristas en Red, "Sistematización."

other organizations, their direct-action work, service provision, and counterdata production challenged dominant norms in the country, contributing, from concrete experiences, to create changes in the legal framework, in particular the 2020 law in Argentina that legalized abortion.[34]

The abortion case exemplifies the broader political nature of missing data. Neither identifying missing data, nor producing counterdata on its own, are sure means of effecting change. This process also necessarily involves political action. As stated in *Counting Feminicide* concerning organizations that produce counterdata on feminicides: "in none of these cases — absolutely zero — do activists think that more data alone can lead to social change."[35] That is, the process of working toward justice does not end with producing or presenting data in either a quantitative or qualitative way. Counterdata production must be part of a broader strategy of engagement and political mobilization. This broader engagement gives data (or its absence) meaning, including criticisms of particular data collection and measurement practices and claims about cases when data should not be collected at all.

Missing Data as a Protection Strategy

Withholding data and protecting it from the reach of the state — effectively producing missing data — can be a protection strategy for minoritized groups. Inequality of power, especially between the state and society, is an important consideration with respect to missing data and strategies of counterdata production that we describe here. Demanding more data and producing more data for the state or by the state is not a guaranteed solution to social problems and, moreover, producing additional data can sometimes

34 Alba Ruibal and Cora Fernandez Anderson, "Legal Obstacles and Social Change: Strategies of the Abortion Rights Movement in Argentina," *Politics, Groups, and Identities* 8, no. 4 (August 2020): 698–713, https://doi.org/10.1080/21565503.2018.1541418.

35 D'Ignazio, *Counting Feminicide*.

actively generate harm.[36] A large body of work demonstrates how minoritized groups — including Black people, queer people, poor people and/or religious minorities — are actively sought out, over-surveilled, and profiled by the state.[37]

For example, Ruha Benjamin discusses databases on gangs in California. Benjamin explains how, once in the database, this population, which is 87 percent Black and/or Latinx, is subjected to increased surveillance and consequently is exposed to more risks, including police violence.[38] Situations like this happen due to historic patterns of racialized policing, spatial exclusion, and state-sanctioned denial of opportunities to racial minorities. However, the harms associated with being counted as data can sometimes be less overt. This is explained in *Data Feminism* with the concept of the "paradox of exposure."[39] The concept describes how the possible gains that might derive from being counted and represented in datasets require being made visible to the state and powerful institutions, which in turn may be dangerous and even deadly for minoritized groups.[40] For example, having more data on LGBTQIA+ people could "inform decisions made about the allocation of resources, changes in legislation, access to services and protections under the law."[41] However, this data might also draw unwanted attention and invite possible targeting and harm. In addition, much data collected about the lives and experiences of these populations often stigmatizes or pathologizes them, marking them as deviant from the normative majority.[42] This form of harm usually results from data collection being undertaken by researchers outside these communities operating in the mode of what is called "dysfunctional rescuing," meaning, "helping based on an assumption that people in the target group cannot help themselves."[43] In these cases, researchers neglect to build relationships with the people who are subject to these datasets.[44]

36 D'Ignazio and Klein, *Data Feminism*.

37 Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, UK: Polity Press, 2019).; D'Ignazio and Klein, *Data Feminism;* *Virginia Eubanks, Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2017); Kevin Guyan, *Queer Data: Using Gender, Sex and Sexuality Data for Action* (London and New York: Bloomsbury Publishing, 2022); Alexis Henshaw, *Digital Frontiers in Gender and Security: Bringing Critical Perspectives Online* (Bristol: Bristol University Press, 2023).

38 Benjamin, *Race After Technology*.

39 D'Ignazio and Klein, *Data Feminism*, p. 71.

40 Os Keyes, "Counting the Countless: Why Data Science is a Profound Threat for Queer People," *Real Life Mag*, April 8, 2019, https://reallifemag.com/counting-the-countless/.

41 Guyan, *Queer Data*.

42 Ibid.

43 Asha Mehta and Kad Smith, "Conflict Resolution with Power and Privilege in Mind," *CompassPoint Nonprofit Services*, 2018, https://www.compasspoint.org/sites/default/files/documents/MANUAL_ConflictResolution_KS-AM_Sep2018.pdf.

44 D'Ignazio and Klein, *Data Feminism*.

This echoes a critique often heard in Indigenous communities with respect to colonial researchers and motivates efforts toward data sovereignty, as described in the "Indigenous Data Sovereignty" entry by Stephanie Russo Carroll, Marisa Duarte, and Max Liboiron.[45] Indigenous groups may also refuse to engage with the colonial state because they do not see it as a legitimate actor and instead assert Indigenous sovereignty over territory and information. For example, in *Counting Feminicide*, Annita Lucchesi recounts how the federal government sought access to her organization's database of Missing and Murdered Indigenous People. With each request, the organization consulted with families who universally felt that the information should *not* be shared. As Lucchesi says, their data production is about "sovereignty and kinship" and not about recognition, reform or cooperation with the state. Thus, from the perspective of the state, which is making the demand, the data is missing.[46]

This example underscores, once again, how missing data is a relational concept — that is, constituted by a demand for data issued from one group or institution to another. In some cases, demands from civil society to the state for missing data may benefit minoritized communities and neglected issues. In other cases, however, minoritized groups may need to actively produce missing data in order to protect themselves from the purview of the state, especially if those datasets could be weaponized in the context of a more extensive configuration of unequal power.[47]

Final Notes

We understand missing data as one part of a larger political process by which different groups make political demands for data and information

45 Stephanie Russo Carrol, Marisa Duarte, and Max Liboiron, "Indigenous Data Sovereignty," in *Keywords of the Datafied State*, eds. Jenna Burrell, Ranjit Singh, and Patrick Davison (Data & Society, 2024).

46 D'Ignazio, *Counting Feminicide*.

47 Ọnụọha, "Missing Data Sets."

to other groups. Civil society may demand data and information from the state or other powerful institutions about key issues. Likewise, the state may demand data from civil society groups who are actively engaged in counterdata production. Additional political engagement is also part of this process and ranges from the production of counterdata to public pressure on the state to collect, maintain, or publish previously missing data, or a combination of those. The demand to acknowledge and take action on different issues can result in practical, political, and public measures from governments, including enacting legislation and implementing public policies, such as providing reparations for families of people who disappeared, developing state policies on gender-related violence, generating official public datasets on environmental issues, and so on. However, political results are not guaranteed from data collection alone. Furthermore, even the missing data concept and its consequent practices can be mobilized from groups with different ideologies; missing data, as we propose here, is fundamentally connected to social justice. As such, addressing structural inequality is the ultimate requirement if the objective is to comprehensively address the challenges associated with an emancipatory missing data perspective and its political and social consequences.

This research was partially funded by the National Science Foundation (award #2213826).

INDIGENOUS DATA SOVEREIGNTY

INDIGENOUS DATA SOVEREIGNTY

BY
STEPHANIE RUSSO CARROLL,
MARISA DUARTE, and MAX LIBOIRON

INDIGENOUS DATA SOVEREIGNTY

By Stephanie Russo Carroll, Marisa Duarte, and Max Liboiron

Introduction

Indigenous Data Sovereignty (IDSov) upholds the rights of Indigenous Peoples, communities, and Nations to "govern the collection, ownership, and application" of datasets created with or about Indigenous communities, Indigenous Lands, and the community's non-human relations.[1] IDSov shifts from Western transactional approaches to data governance characterized by rights-based, relational approaches that enact responsibilities to Peoples and Land. We describe IDSov as a means to disrupt colonial infrastructures, policies, and practices through centering Indigenous systems of governance and knowledges. Indigenous Peoples have diverse, specific relationships within their Lands, so there is no single approach to IDSov.[2] We might best think of IDSov as a social movement that began by Indigenous Peoples in the 1990s living within English-speaking settler-colonial nation-states (see the principles section below), so while IDSov might claim examples and practitioners, not all cases and practitioners use the term or identify with IDSov. Ultimately, as Elders remind us, IDSov renews our ancestral instructions — our traditions, protocols, and responsibilities for the care and transmission of communal knowledges and information — in the digital world.

1 Stephanie Russo Carroll, Desi Rodriguez-Lonebear, and Andrew Martinez, "Indigenous Data Governance: Strategies from United States Native Nations," *Data Science Journal* 18, no. 1 (July 2019): 31, https://doi.org/10.5334/dsj-2019-031; Tahu Kukutai and John Taylor, *Indigenous Data Sovereignty: Toward an Agenda* (Canberra: ANU Press, 2016).

2 Kukutai and Taylor, *Indigenous Data Sovereignty.*

A key concept within IDSov is Indigenous sovereignty itself. In the Americas, Aotearoa, Australia and other places, European colonizers signed treaties with Indigenous Peoples to establish boundaries for distinct neighboring governments. Indigenous Peoples signed with the force of their inherent sovereignty: the political will of Peoples who know those Lands and territories to be rightfully their own. However, European colonizers signed such treaties with the goals of settlement and national expansion. Leaders and citizens within modern nation-states have come to confuse self-governing Indigenous Peoples for minority populations within the settler nation-state (the government put in place by colonists that survives today). This is a settler mentality, one that: a) consciously and unconsciously legislates that Indigenous Peoples either no longer exist or do not have a legal right to self-govern, and b) presumes that the modern, settler nation-state is the ideal mode of governance for managing the affairs of Indigenous Peoples. Any analysis of data governance that assumes the settler state is the sole sovereign for making data-driven decisions also marginalizes Indigenous Nations and our data practices, ethics, and infrastructures. IDSov responds to these assumptions.

IDSov is premised on Indigenous sovereignty and our continued governance of our Lands and Peoples. In data-sharing protocols, Indigenous Peoples are not stakeholders or interest groups such as industry partners or NGOs. Instead, Indigenous Peoples bear distinct *legal and moral rights* that supersede commercial interests. We are *rights-holders*, not stakeholders. Accordingly, IDSov research and data protocols defend a broad range of Indigenous rights.[3] Many of these rights are outlined within the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). For example, the right to practice and revitalize cultural traditions and customs: "this includes the right to maintain, protect, and develop the past, present,

3 Maui Hudson, Stephanie Russo Carroll, Jane Anderson, Darrah Blackwater, Felina Cordova-Marks, Jewel Cummins, Domonique David-Chavez, Adam Fernandez, Ibrahim Garba, Danielle Hiraldo, Mary Beth Jager, Lydia L. Jennings, Andrew Martinez, Rogena Sterling, Jennifer D. Walker, and Robyn Rowe, "Indigenous Peoples' Rights in Data: A Contribution Toward Indigenous Research Sovereignty," *Frontiers in Research Metrics and Analytics* 8 (May 2023), https://www.frontiersin.org/articles/10.3389/frma.2023.1173805/full.

and future manifestations of their cultures, such as archaeological and historical sites, artefacts, designs, ceremonies, technologies, and visual and performing arts and literature."[4]

Designing data-sharing protocols with regard to UNDRIP and IDSov results in: Indigenous-centered practices of provenance (origin) and attribution (authorship) in labeling datasets;[5] requirements to abide by sovereign Indigenous jurisdiction and law;[6] informed consent, including collective community consent;[7] limiting data collection with regard for sacred spaces, seasons, situations, and sub-populations;[8] and monitoring data sharing to ensure appropriateness, accuracy, meaningful use, confidentiality, and the overall security of the sovereign nation.[9]

IDSov work reveals how Indigenous Peoples' approaches to data collection, data sharing, and data governance are not about retrofitting settler-state forms of data governance for Indigenous communities and Nations. As a mode of governance, Indigenous sovereignty is generally not founded on the presumption of private property, generating profit through alienating labor from the land, and the accumulation of wealth toward personal happiness. Indigenous sovereignty is rather a paradigm of governance premised on obligations to Land, relatives (humans and not), future generations, and ancestors. For example, if someone obtains data within an IDSov paradigm, the question is not "what can I do with this data?" but "to whom am I obliged with this data? What does this data and its data holder owe to community and Land, and how do I best meet those obligations in how this data is stored, shared (or not) and interpreted?"[10] In an interview with Northern Cheyenne demographer Desi Small-Rodriguez, an Elder said, "Sovereignty as tribal nations was given to us by the Creator. It is sacred.

4 Megan Davis, "Data and the United Nations Declaration on the Rights of Indigenous Peoples," in *Indigenous Data Sovereignty*, eds. Tahu Kukutai and John Taylor (Canberra: ANU Press, 2016); The United Nations, "United Nations Declaration of the Rights of Indigenous Peoples," https://social.desa.un.org/issues/indigenous-peoples/united-nations-declaration-on-the-rights-of-indigenous-peoples.

5 Jane Anderson and Kimberly Christen, "Decolonizing Attribution: Traditions of Exclusion," *Journal of Radical Librarianship* 5 (June 2019): 113–52.

6 William Haney, "Protecting Tribal Skies: Why Indian Tribes Possess the Sovereign Authority to Regulate Tribal Airspace," *American Indian Law Review* 40, no. 1 (January 2016): 1; Sarah D. Littletree, "'Let Me Tell You About Indian Libraries': Self-Determination, Leadership, and Vision," (PhD diss., University of Washington, 2018).

7 Krystal Tsosie, Joe Yracheta, and Donna Dickenson, "Overvaluing Individual Consent Ignores Risks to Tribal Participants," *Nature Reviews Genetics* 20 (July 2019): 1, https://doi.org/10.1038/s41576-019-0161-z; Max Liboiron, Alex Zahara, and Ignace Schoot, "Community Peer Review: A Method to Bring Consent and Self-Determination into the Sciences," *Preprints* (June 2018), https://doi.org/10.20944/preprints201806.0104.v1.

8 Kawika B. Winter, Noa Kekuewa Lincoln, Fikret Berkes, Rosamma A. Alegado, Natalie Kurashima, Kiana L. Frank, Pua'la Pascua, Yoshini M. Rii, Frederick Reppen, Ingrid S.S. Knapp, Will C. McClatchey, Tamara Ticktin, Celia Smith, Erik c. Franklin, Kristen Oleson, Melissa R. Price, Margaret A McManus, Megan J. Donahue, Kuulei S. rodgers, Brian W. Bowen, Craig E. Nelson, Bill Thomas, Jo-Ann Leong, Elizabeth M.P Madin, Malia Ana J. Rivera, Kim A. Falinski, Leah L. Bremer, Jonathan L. Deenik, Sam M. Gon III, Bran Neilson, Ryan Okano, Anthony Olegario, Ben Nyberg, A. Hijei Kawelo, Kelij Kotubetey, J. Kanekoa Kukea-Shultz, and Robert J. Toonen, "Ecomimicry in Indigenous Resource Management: Optimizing Ecosystem Services to Achieve Resource Abundance, with Examples from Hawaii," *Ecology and Society* 25, no. 2 (2020): 26, https://doi.org/10.5751/ES-11539-250226.

9 Diane E. Smith, "Governing Data and Data for Governance: The Everyday Practice of Indigenous Sovereignty," in *Indigenous Data Sovereignty*, eds. Tahu Kukutai and John Taylor (Canberra: ANU Press, 2016).

10 Jennifer Wemigwans, *A Digital Bundle: Protecting and Promoting Indigenous Knowledge Online* (Regina: University of Regina, 2018).

Data to exercise our sovereignty is also sacred."[11] This means that IDSov is not simply about Indigenous individuals collecting data toward an imagined gain, but rather is also a form of Indigenous *governance* through data based *in right relation*. It also means that under specific circumstances, because of the relational requirements, non-Indigenous people can be part of but never solely responsible for IDSov work.

Colonial Legacies in Data Infrastructures and Practices and Our Responses

For centuries, settler-colonial and colonial states have used data to control, erase, and enact genocide against Indigenous Peoples, from military intelligence to national censuses[12] to standardized education tests.[13] Data practices around Indigenous Peoples continue to benefit non-Indigenous Peoples today. For example, Inuit Tapiriit Kanatami, the national representational organization for Inuit in Canada, writes that:

> Inuit in Canada are among the most studied Indigenous peoples on earth. The primary beneficiaries of Inuit Nunangat [Inuit homelands] research continue to be researchers themselves, in the form of access to funding, data and information, research outcomes, and career advancement. Inuit remain largely marginalized from research governing bodies and in turn from experiencing the benefits of research.[14]

As another example, Indigenous scholars have commented on how intellectual property regimes are designed toward commodification of Indigenous Knowledge (IK), contributing to promiscuous uses of datasets and information about Indigenous Peoples.[15] When combined with state

11 Desi Rodriguez-Lonebear, "Building a Data Revolution in Indian Country," in *Indigenous Data Sovereignty*, eds. Tahu Kukutai and John Taylor (Canberra: ANU Press, 2016).

12 Chris Andersen, "From Nation to Population: The Racialisation of 'Métis' in the Canadian Census," *Nations and Nationalism* 14, no. 2 (2008): 347–68, https://doi.org/10.1111/j.1469-8129.2008.00331.x.

13 Wendy M. Pearce and Cori Williams, "The Cultural Appropriateness and Diagnostic Usefulness of Standardized Language Assessments for Indigenous Australian Children," *International Journal of Speech-Language Pathology* 15, no. 4 (August 2013): 429–40, https://doi.org/10.3109/17549507.2012.762043.

14 Wemigwans, *A Digital Bundle*, 5; Inuit Tapiriit Kanatami, *National Inuit Strategy on Research* (Ottawa: Inuit Tapiriit Kanatami, 2018), https://www.itk.ca/wp-content/uploads/2018/04/ITK_NISR-Report_English_low_res.pdf

15 Pinar Oruç, "Documenting Indigenous Oral Traditions: Copyright for Control," *International Journal of Cultural Property* 29, no 3 (2022), 243–64. doi:10.1017/S0940739122000273.

practices of record keeping sans Indigenous Nation-to-Nation trust negotiations, such uses are unearned, transactional, and sustaining conditions of elite cultural theft, appropriation, resource extraction, and data violence against Indigenous Peoples.[16] Indigenous Peoples have found that measures from settler perspectives based in good intentions and respect for "all parties" often
amplify rather than mitigate existing power asymmetries. For example, calls for open data, public reporting, and transparency often do not address colonial politics and the structural inequities shaping Indigenous marginalization, resulting in more non-Indigenous access to Indigenous data.[17]

In response, much IDSov practice deliberately disrupts discrimination against Indigenous Peoples from the data practices of the settler state. For example, Indigenous Nations deal with the categorizations of Indigeneity via "blood quantum" (a settler-state concept) by making their own policies for who belongs to their communities and how to record that data, or challenge how the settler state's census fails to capture meaningful data about Indigenous Peoples.[18] Indigenous Peoples may offer their own data collection and interpretation. These strategically align with and oppose aspects of colonial scientific paradigms in, for example, archaeology[19] or environmental contamination,[20] both of which continue traditions of genocide, Indigenous erasure, Land-based harm, and non-Indigenous access to Indigenous Lands.[21]

Often Indigenous researchers develop community- and discipline-specific methods of control that prevent misuse of datasets, in particular where misuse could result in unjust criminalization of activists, resource extraction, or misinterpretation in courts of law. For example, a requirement

16 Sue McKemmish, Livia Iacovino, Eric Ketelaar, Melissa Castan, and Lynette Russell, "Resetting Relationships: Archives and Indigenous Human Rights in Australia," *Archives and Manuscripts* 39 (2011): 107–144.

17 Walter et al., *Indigenous Data Sovereignty*; Stephanie Carroll Rainie, Tahu Kukutai, Maggie Walter, Oscar Lusi Figueroa-Rodriguez, Jennifer Walker, and Per Axelsson, "Indigenous Data Sovereignty," in *The State of Open Data: Histories and Horizons*, eds. Tim Davies, Stephen B. Walker, Mor Rubinstein, and Fernando Perini (Cape Town: African Minds, 2019), 300–320.

18 Desi Rodriguez-Lonebear, "The Blood Line: Racialized Boundary Making and Citizenship among Native Nations," *Sociology of Race and Ethnicity* 7, no. 4 (October 2021): 527–42, https://doi.org/10.1177/2332649220981589; Julie Wailing, Desi Small-Rodriguez, and Tahu Kukutai, "Tallying Tribes: Waikato-Tainui in the Census and Iwi Register - Ministry of Social Development," *Social Policy Journal of New Zealand*, no. 36 (2009), https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/journals-and-magazines/social-policy-journal/spj36/36-tallying-tribes.html.

19 Matthew C. Sanger and Kristen Barnett, "Remote Sensing and Indigenous Communities: Challenges and Opportunities," *Advances in Archaeological Practice* 9, no. 3 (August 2021): 194–201, https://doi.org/10.1017/aap.2021.19.

20 Michelle Murphy, "Alterlife and Decolonial Chemical Relations," *Cultural Anthropology* 32, no 4 (2017): 494–503, https://doi.org/10.14506/ca32.4.02.

21 Max Liboiron, *Pollution is Colonialism* (Durham: Duke University Press, 2021).

in Canada's Species at Risk Act has specific language about the inclusion of Indigenous Traditional Knowledge when determining the risk or recovery of species.[22] In some ways, this is a good thing, but there are also concerns that the inclusion of Traditional Knowledge in settler-state policy-driven documentation means that sacred, protected, private, place-based, and relational knowledge can be discoverable in court and used in ways that are not appropriate or consented to. In response, Indigenous Peoples use a variety of techniques to control our data, such as encryption, timed destruction of datasets and keys, anonymization, selective reporting, access limitations, metadata that relay protocols and permissions, use of Indigenous languages or dialects, and reliance on tribal regulations, such as research review processes. Indigenous communities, organizations, and governments also informally educate community members and outsiders about data privacy and security, personal information management, research quality, and informed consent toward protecting the Nation. Indeed, much IDSov work concerns the creation, ownership, control, access, possession, and demonstration of collective benefits around scientific data practices, including Indigenizing infrastructure and data regimes, building trustworthy relationships, and addressing planned and future use of datasets and even data infrastructure.[23]

IDSov as Responsibility to Land

IDSov is a responsibility to Land. Here, Land doesn't refer to just dirt and bees and trees (though it includes those too), but also waters, stars, histories, spirit, ancestors, future ancestors, and place (each variously defined by different Indigenous cosmologies). IDSov is rooted in the roles and responsibilities for caretaking the systems of knowledge that Indigenous Peoples have relied on since time immemorial as they have lived in relation to Land.

22 "Fisheries and Oceans Canada Species at Risk Act Listing Policy and Directive for 'Do Not List' Advice," Fisheries and Oceans Canada, 2013, https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/365882.pdf.

23 Stephanie Russo Carroll, Edit Herzcog, Maui Hudso, Keith Russell, and Shelley Stall, "Operationalizing the CARE and FAIR Principles for Indigenous Data Futures," *Scientific Data 8*, no. 1 (April 2021): 108, https://doi.org/10.1038/s41597-021-00892-0; Marisa Elena Duarte, "Native and Indigenous Women's Cyber-Defense of Lands and Peoples," in *Networked Feminisms: Activist Assemblies and Digital Practices*, eds. Shana MacDonald, Brianna I. Wiens, Michelle MacArthur, and Milena Radzikowska (London: Lexington Books, 2022); *CARE Principles for Indigenous Data Governance* The Global Indigenous Data Alliance, accessed May 5, 2023, https://www.gida-global.org/care.

As such, IDSov is place-based and Nation-specific, rather than universal and amenable to standardization. For example, rules within the Pueblos of New Mexico prohibit taking and disseminating photos or videos of ceremonies and events. This rule is designed to uphold group privacy in a state that profits from a romanticized image of Pueblo Indigeneity. But other Indigenous Peoples do not have this rule. Another example includes the commitment of the nonprofit Village Earth to generate many data dashboards that summarize data about Indigenous Peoples and Lands through their intertribal Native Land Information System.[24] Approaches vary across the Indigenous pluriverse, especially as differently positioned Indigenous Peoples either do not have the same type of data or caution against sharing it openly.[25]

IDSov approaches can also vary by academic discipline and field. For example, due to the nature of consent over human tissues, Indigenous genomics researchers assert a "DNA on loan" standard instead of "gifting" tissues for precision medicine research because the disciplinary standard of individual consent conflicts with IDSov principles of collective control.[26]

Indigenous legal scholars assert sovereign rights to regulate airspace and airwaves for sciences dependent on data from satellites, drones, spectrum, and wide area networks that can access Indigenous Land without setting foot on it.[27] Health researchers often work with accumulated datasets of many people, making it difficult to identify a single governing council who can supervise IDSov principles of reciprocity and accountability. In one case, health researchers report the practical limitations of ownership, control, access, and possession (OCAP) standards for studies indexing end-of-life care among an urban Indigenous patient group in Canada, and thus developed an alternative standard for a regional palliative care provider.[28]

24 "Native Land Information System," Native Land Information System, accessed May 5, 2023, https://nativeland.info/.

25 Walter et al., "Indigenous Data Sovereignty."; Kimberly A. Christen, "Does Information Really Want to Be Free? Indigenous Knowledge Systems and the Question of Openness," *International Journal of Communication* 6 (November 2012): 24.

26 Krystal S. Tsosie, Joseph M Yracheta, Jessica A. Kolopenuk, and Janis Geary, "We Have 'Gifted' Enough: Indigenous Genomic Data Sovereignty in Precision Medicine," *The American Journal of Bioethics* 21, no. 4 (April 2021): 72–75, https://doi.org/10.1080/15265161.2021.1891347.

27 Haney, "Protecting Tribal Skies"; M.L. Cornette and B.L. Smith, "Electronic Smoke Signals: Native American Radio in the United States," *Cultural Survival Quarterly* 22 (1998): 28–31.

28 Sarah Funnell, Peter Tanuseputro, Angeline Letendre, Lisa Bourque Bearskin, and Jennifer Walker, "'Nothing About Us, Without Us.' How Community-Based Participatory Research Methods Were Adapted in an Indigenous End-of-Life Study Using Previously Collected Data," *Canadian Journal on Aging* 39, no. 2 (June 2020): 145–55, https://doi.org/10.1017/S0714980819000291.

Understanding how IDSov varies in practice and theory across disciplines and fields demonstrates the mutability of data practices and the ethical limits of scientific disciplines as they approach the territorial, sovereignty, and autonomy concerns of diverse Indigenous Peoples.

A responsibility to Land is what distinguishes IDSov approaches to data and information sharing (including concepts like digital bundles) from other justice-oriented data efforts such as open access, digitization for the public good, digital archiving, digital storytelling, and individual consent.[29] Even merely being an Indigenous person gathering data does not alone address IDSov because of what Indigenous sovereignty means: our right to exercise our collective responsibilities to Land through governance.

IDSov Is Related to Indigenous Knowledge Work

IDSov overlaps with IKs and traditional ecological knowledge (TEK) work, but they are not synonymous. In the 1980s, Indigenous scholars began seeking a term to reflect Indigenous ways of knowing. They began adapting "Indigenous Knowledge" — an anthropological term — toward the goals of Indigenous intellectual autonomy. TEK was similarly debated and developed.[30] From an Indigenous perspective, IKs are intergenerational ancestral systems of knowledge that reflect "Indigenous informed epistemologies" through place-based dimensions of tradition, empiricism, and revelation or insight.[31] Due to the social positions of Indigenous Peoples, IKs are inherently decolonial or anti-colonial.

Indigenous systems of knowledge are not data, but rather a system of relations over time. Indigenous systems of knowledge are safeguarded by sanctioned Indigenous Peoples who sustain them through self-governance, philosophy, language, medicine, science, and ceremony. As such, IK cannot

29 Christen, "Does Information Really Want to be Free?"; Wemigwans, *A Digital Bundle*.

30 Charles Kamau Maina, "Power Relations in the Traditional Knowledge Debate: A Critical Analysis of Forums," *International Journal of Cultural Property* 18, no. 2 (2011): 143–78, https://doi.org/10.1017/S0940739111000130; Laurence Helfer and Graeme Austin, "Indigenous Peoples' Rights and Intellectual Property," in *Human Rights and Intellectual Property: Mapping the Global Interface*, eds. Laurence Helfer and Graeme W. Austin (Cambridge: Cambridge University Press, 2011), 432–502.

31 George J. Sefa Dei, "Rethinking the Role of Indigenous Knowledges in the Academy," *International Journal of Inclusive Education* 4, no. 2 (April 2000): 111–32, https://doi.org/10.1080/136031100284849.

be divorced from knowledge holders, community, protocol, and obligation. Indeed, IK gains both meaning and applicability precisely through the ancestral community-esteemed protocols of sharing governing its practice, even when such sharing occurs via digital bundles in online spaces.[32] This is why, at its basis, IDSov work, originating in Indigenous systems of knowledge and multiplying legal protections for IK and TEK, occurs through Nation-to-Nation agreements regarding Indigenous Peoples and institutions.

Most importantly, IK does not and cannot fit in a spreadsheet, even though IK-centered methods can produce spreadsheets in technique.[33] Divorcing IK from the knowledge-keepers, landscapes, languages, telling, aurality, and philosophies of its emergence is to designify the knowing, thus committing scientific extraction.[34] Indigenous Peoples often face well-intentioned inclusion models that insist on IK without people, archives without protocol, and storytelling without obligation.[35] This is and can only be expropriation, if not appropriation. In these cases, inclusion opposes the principles of IDSov.

There have been many calls and efforts to include IKs and TEK in studies conducted by non-Indigenous researchers. Yet without a real-world understanding of Land relations, including what Leanne Betasamosake Simpson (Anishinaabe) calls "constellations of coresistance," no amount of or care for Indigenous data, information, or stories in research can support an Indigenous way of knowing or observation.[36] Moreover, positioning IK and TEK as data, information, or anecdotes ripens fragments of knowledge for exploitation. This point is maintained by many Indigenous TEK practitioners, including Tribal historic preservation officers who must often, on behalf of their Tribal government, satisfy Western scientific demands through translating Tribal ways of knowing and decision-making into datafied Western structures.

32 Wemigwans, *A Digital Bundle*.

33 Maggie Walter and Chris Andersen, *Indigenous Statistics: A Quantitative Research Methodology* (Walnut Creek: Left Coast Press, 2013).

34 Dylan Robinson, *Hungry Listening: Resonant Theory for Indigenous Sound Studies* (Minneapolis: University of Minnesota Press, 2020); Julie Cruikshank, *Do Glaciers Listen?: Local Knowledge, Colonial Encounters, and Social Imagination* (Vancouver: UBC Press, 2010).

35 Sandra Littletree, Miranda Belarde-Lewis, and Marisa Duarte, "Centering Relationality: A Conceptual Model to Advance Indigenous Knowledge Organization Practices," *Knowledge Organization* 47, no. 5 (November 2020): 410–426, https://digital.lib.washington.edu:443/researchworks/handle/1773/46601; Ricardo L. Punzalan and Michelle Caswell, "Critical Directions for Archival Approaches to Social Justice," *The Library Quarterly* 86, no. 1 (January 2016): 25–42, https://doi.org/10.1086/684145.

36 Leanne Betasamosake Simpson, "Constellations of Coresistance," in *As We Have Always Done: Indigenous Freedom Through Radical Resistance*, ed. Leanne Betasamosake Simpson (Minneapolis: University of Minnesota Press, 2017).

Yet data and information are not the opposite of IK. Indigenous practitioners of informatics — including data scientists, computer scientists, social scientists, library and information scientists, geneticists, and environmental scientists — foreground Indigenous relationships with each other and with Indigenous Lands in their own concepts and creation of data, information, and knowledge, thereby developing methodologies and advancing Indigenous sciences.[37] For instance, some Indigenous statisticians are using Indigenous community priorities, ethics, cosmologies, and numeracy traditions to guide their interpretation of data.[38] IDSov and Indigenous data regulation offer means to ethically create, protect, and control datasets that are outcomes of IK and TEK.

Frameworks and Principles

In the 2010s, Indigenous researchers (mostly empiricists) from Canada, New Zealand/Aotearoa, the United States, and Australia (CANZUS) developed the overarching principles of IDSov. These researchers bear strong commitments to the principles of Indigenous science and self-determination. Many Indigenous Peoples of the CANZUS countries bear treaty or other sovereign recognition relationships with the dominant settler nation-state. Country-level discussions about IDSov are often coordinated through hubs in each country, including the First Nations Information Governance Center (FNIGC) in Canada, the Te Mana Rauranga Māori Data Sovereignty Network in Aeoteroa, the US Indigenous Data Sovereignty Network, and Maiamnayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Network.[39]

37 Littletree et al., "Centering Relationality"; Jessie Loyer, "Collections Are Our Relatives Disrupting the Singular, White Man's Joy That Shaped Collections," in *The Collector and the Collected: Decolonizing Area Studies Librarianship*, eds. Megan Browndorf, Erin Pappas, and Anna Arays (Sacramento: Library Juice Press, 2021); Liboiron, *Pollution is Capitalism*.

38 Walter and Andersen, *Indigenous Statistics;* Ella Henry and C. Crothers, *Exploring Papakāinga: A Kaupapa Māori Quantitative Methodology* (Porirua: National Science Challenges Building Better Homes, Towns and Cities Ko ngā wā kāinga hei whakamāhorahora, 2019).

39 Ray Lovett, Vanessa Lee, Tahu Kukutai, Stephanie Carroll Rainie, Jennifer Walker, "Good Data Practices for Indigenous Data Sovereignty," in *Good Data*, eds. Angela Daly, Kate Devitt, and Monique Mann (Amsterdam: Institute of Network Cultures, 2019).

OCAP was developed in 2009 through Canada's Assembly of First Nations Chiefs-in-Assembly, which subsequently became FNIGC. FNIGC trademarked the framework in 2015 to "assert that First Nations have control over data collection processes, and that they own and control how this information can be used."[40] Seeking trademark protection emerged from the need to enforce the accuracy and purpose of the framework, especially as non-Indigenous researchers began distorting the acronym and its terms, and to leverage an array of protections for research datasets and information bearing specifically First Nations access and use rights. While inspiring to others, OCAP is specifically for First Nations contexts. The OCAP framework[41] includes the following principles:

- "Ownership refers to the relationship of First Nations to their cultural knowledge, data, and information. This principle states that a community or group owns information collectively in the same way that an individual owns his or her personal information.
- "Control affirms that First Nations, their communities, and representative bodies are within their rights to seek control over all aspects of research and information management processes that impact them. First Nations control of research can include all stages of a particular research project-from start to finish. The principle extends to the control of resources and review processes, the planning process, management of the information and so on.
- "Access refers to the fact that First Nations must have access to information and data about themselves and their communities regardless of where it is held. The principle of access also refers to the right of First Nations' communities and organizations to manage and make decisions regarding access to their collective information."

40 "The First Nations Principles of OCAP," First Nations Information Governance Centre, accessed February 24, 2023, https://fnigc.ca/ocap-training/.

41 "The First Nations Principles of OCAP," First Nations Information Governance Centre.

This may be achieved, in practice, through standardized, formal protocols.

- "Possession: While ownership identifies the relationship between a people and their information in principle, possession or stewardship is more concrete: it refers to the physical control of data. Possession is the mechanism by which ownership can be asserted and protected."

Similarly, in 2015, groups of Māori researchers and technology specialists created the concept for a Māori data sovereignty network during a presentation on IDSov in Canberra, Australia.[42] At gatherings in Aotearoa in 2015 and 2016, they developed an IDSov charter. The Te Mana Rauranga Charter specifies rights within the Treaty of Waitangi and UNDRIP, and bridges technoprogressive ideologies of "data as world of opportunity" with time-tested principles of whanaungatanga, rangatiratanga, and kotahitanga toward governance; and manaakitanga, kaitiakitanga, and whakapapa toward operations. Each principle is carefully considered with regard to technical infrastructures, private industry roles, research praxis, and novel innovations within the information economies in Aotearoa/New Zealand. Understanding the specificity of the regional and state-based frameworks is thus integral for researchers developing reciprocal and responsible data-driven research or industry relationships with specific Indigenous partners. The Te Mana Rauranga Charter[43] includes:

- Whanaungatanga and Whakapapa: "in Māori thinking and philosophy relationships between [humans], Te Ao Turoa (the natural world), and spiritual powers inherent therein, and Taha Wairua (spirit) are everything. Whakapapa evidences those linkages."

42 "Māori Data Sovereignty Network Charter," Te Mana Raraunga, https://www.temanararaunga.maori.nz/tutohinga.

43 "Māori Data Sovereignty."

- Rangatiratanga: the "iwi/Māori aspiration for self-determination, to be in control of our own affairs and to influence those taking place within our iwi boundaries … Rangatiratanga can be expressed through leadership and participation."
- Kotahitanga: relationality based on "a collective vision and unity of purpose while recognising the mana of rangatira from individual hapū and iwi."
- Manaakitanga: "the responsibility to provide hospitality and protection to whānau, hapū, iwi, the community, and the environment. The foundations of manaakitanga rely on the ability of Māori to live as Māori, to access quality education, to have good health, to have employment opportunities and to have liveable incomes."
- Kaitiakitanga: being "an effective steward or guardian and relates to actions that ensure a sustainable future for all people."

In 2019 at a conference in Basque territory led by statistician Maggie Walters (Palawa) and demographer Desi Small-Rodriguez (Northern Cheyenne), participants formed the Global Indigenous Data Alliance (GIDA). At that meeting, GIDA members approved relationships and rights discourses for asserting a balance across First Nations principles of ownership, control, access, protection, and reciprocity (OCAP principles) toward broader CARE Principles for Indigenous Data Governance — collective benefit, authority to control, responsibility, and ethics. GIDA members examine how these merge with the FAIR Principles of findability, accessibility, interoperability, and reusability.[44] GIDA has also translated the IDSov discourse

44 Stephanie Russo Carroll, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, Kay Raseroka, Desi Rodriguez-Lonebear, Robyn Rowe, Rodrigo Sara, Jennifer D. Walker, Jane Anderson, and Maui Hudson, "The CARE Principles for Indigenous Data Governance," *Research Data Alliance* 19, (2020), DOI: 10.5334/dsj-2020-043; Stephanie Russo Carroll et al., "Operationalizing the CARE and FAIR Principles"; Neha Gupta, Andrew Martindale, Kisha Supernant, and Michael Elvidge, "The CARE Principles and the Reuse, Sharing, and Curation of Indigenous Data in Canadian Archaeology," *Advances in Archaeological Practice* 11, no. 1 (February 2023): 76–89, https://doi.org/10.1017/aap.2022.33.

into Spanish, Vietnamese, German, and Khmer (it was already born in part through discussions in Te Reo Māori) toward inclusivity.[45] The CARE Principles are:

- "Collective Benefit. Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data. C1) For inclusive development and innovation. C2) For improved governance and citizen engagement. C3) For equitable outcomes.
- "Authority to Control. Indigenous Peoples' rights and interests in Indigenous data must be recognized and their authority to control such data respected. A1) Recognizing rights and interests. A2) Data for governance. A3) Governance of data.
- "Responsibility. Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit. R1) For positive relationships. R2) For expanding capability and capacity. R3) For Indigenous languages and worldviews.
- "Ethics. Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem. E1) For minimizing harm and maximizing benefit. E2) For justice. E3) For future use."

It would be impossible for these frameworks to be adopted as-is in other contexts, particularly non-Indigenous movements or struggles. The specificity of the ways Indigenous Peoples defend their rights to data-driven practices relating to their Peoples, Land, governments, and economies is apparent in the range of local protections, from codes and laws integrating the latest IDSov principles to pre-existing research regulations to customary practices

45 "CARE Principles for Indigenous Data Governance," The Global Indigenous Data Alliance.

46 Wemigwans, *A Digital Bundle*.

47 Rodriguez-Lonebear, "Building a Data Revolution in Indian Country."

and structures that restrict expropriation. This is perhaps the most important foundation of IDSov: that particular Indigenous Peoples in various parts of the world have place-based strategies toward data-driven self-defense and continuance. We thus describe these various principles in their own terms; generalizing or synthesizing them without an anchor to their places runs counter to the spirit in which they were created. As such, Indigenous researchers and other data actors working in other nations must be conscientious that they are not importing principles from another context, or misinterpreting principles, in particular as these are deeply rooted in ancestral Indigenous philosophies.

The Near Future of IDSov

Indigenous intellectualism, Indigenous science, Indigenous ways of knowing, the right to know, and IK/TEK debates have existed for centuries prior to the IDSov movement. As a relatively new movement and discourse fed by many wellsprings, IDSov is emergent. Some aspects of IDSov have yet to be clarified. Some of these relate to the question of privacy, and how Indigenous privacy and security contexts reveal the technical and social erosion of legal privacy worldwide through reliance on third-party information sharing for industrial computation. For example, US Tribes may realize the ways IDSov conflicts with law enforcement data sharing in particular as the Supreme Court challenges sovereign jurisdictions, including surrounding child welfare cases and missing persons cases. IDSov's promise is dependent on the capacity of self-governing Indigenous Peoples to assemble and sustain a data management ecosystem sufficiently robust to keep pace with governments, universities, and industry. IDSov is also dependent on an assumption that the participating Indigenous governments support some kind

of democratic approach to research and data, or, alternatively, do not regulate research toward censoring their own Peoples.

As Indigenous Peoples we must create our own data practices and strategize how — and whether — to work with settler-state data practices and non-Indigenous researchers in ways that promulgate Indigenous goals and priorities. IDSov must emerge through the relationship between Indigenous Peoples and our governments, beginning whenever possible through Indigenous-Nation-to-Indigenous-Nation agreements. This allows online datasets, information, or bodies of knowledge that are intentionally created to be passed on as a matter of tradition in ways that are meant to be interpreted as interconnected, "aspects of long-existing Indigenous Knowledge in new formats and in relation to new contexts," or as Wemigwans has described them, "digital bundles."[46] In this sense, IDSov work is about protecting what is already ours as Indigenous Peoples, even across new formats and emerging sociotechnical relationships. Because of this spirit of resurgence, Rodriguez-Lonebear has characterized IDSov work as a data revolution.[47]

There is another kind of challenge for Indigenous Peoples who agree with IDSov principles, but who, due to absence of legal personhood or rights to sovereignty, lack a means of enforcement. This is the case for many Indigenous Peoples throughout Abiayala and in states throughout the world where there is no legal path to recognition or where corruption is pervasive and persistent. These types of challenges will likely have to be fought via human rights cases through courts that rely on international pressure to right wrongs. As such, a related challenge pertains to the work of transnational learning that could occur through specialists who can compare and translate aspects of other types of data justice movements in other countries and

contexts. Though, due to the differences in legal structures, it is unlikely that there will ever be a supranational set of codes across, say, the European data justice movements and IDSov movements, such discussions nevertheless encourage healthy understanding about the ways data-driven practices intersect with human rights, anti-oppression, healthy economies, digital equity, and technological futures among many peoples worldwide. In such a spirit of conversation, of sharing knowledge and experience, GIDA exemplifies the assertion that while IDSov practices are not universal, they are pluriversal, and align through a common struggle against extractive and weaponized colonial knowledge practices. A key contribution of IDSov is the premise that Indigenous self-determination is a sound means of defense, and on this basis, we light our fires for the strength and flourishing of our Indigenous Peoples, Lands, friends, and accomplices worldwide.

FURTHER
FURTHER
FURTHER

READINGS
READINGS
READINGS

Further Readings

Arranged chronologically, these resources inform the broader conception and ongoing engagement with the datafied state on a global scale.

1.	Abrams, Philip. 1988. "Notes on the Difficulty of Studying the State (1977)." *Journal of Historical Sociology* 1 (1): 58–89. https://doi.org/10.1111/j.1467-6443.1988.tb00004.x.	Juxtaposing state as a system and state as an idea, this paper is an effort to demystify the nature of the state as an object of inquiry and an invitation to generative possibilities of analysis when we consider that it does not exist.
2.	Sharma, Aradhana., and Akhil Gupta, eds. 2006. *The Anthropology of the State: A Reader. Malden,* MA: Blackwell Publishing.	This reader provides an innovative combination of classic social theory texts and ethnographic case studies to offer an anthropological framework for studying institutions, practices, and experiences that underlie the cultural constitution of the state.
3.	Holston, James. 2007. *Insurgent Citizenship: Disjunctions of Democracy and Modernity in Brazil.* Princeton: Princeton University Press.	Mapping the struggles of Brazilians in São Paulo's peripheries to secure effective legal title over their homes, this book calls attention to asymmetrical forms of citizenship that emerge in the making and management of difference in accessing city infrastructure.
4.	Walter, Maggie, and Chris Andersen. 2013. *Indigenous Statistics: A Quantitative Research Methodology.* Walnut Creek: Left Coast Press.	Using examples of research projects from First World Indigenous peoples in the United States, Australia, and Canada, this book explores how quantitative methods and indigenous ways of knowing can mutually shape each other.

5.	Bierschenk, Thomas, and Jean-Pierre Olivier de Sardan, eds. 2014. *States at Work: Dynamics of African Bureaucracies. Africa-Europe Group for Interdisciplinary Studies, volume 12.* Leiden ; Boston: Brill.	This book draws much needed empirical attention to processes of state-building and public bureaucracies in African countries by focusing on everyday lives of street-level bureaucrats, routines of public service delivery, and ordinary interventions in organizing public administration reform.
6.	Mazzucato, Mariana. 2015. *The Entrepreneurial State: Debunking Public vs. Private Sector Myths.* New York: PublicAffairs.	This book positions the state as a primary risk-taker in fostering entrepreneurial innovation and makes a compelling case for inequality in the distribution of rewards of such risk-taking between the state and the private sector.
7.	Rottenburg, Richard, Sally E. Merry, Sung-Joon Park, and Johanna Mugler, eds. 2015. *The World of Indicators: The Making of Governmental Knowledge through Quantification.* Cambridge, United Kingdom: Cambridge University Press.	Making an intervention in quantitative efforts to govern social life, the edited volume examines the politics of indicators and indices as proxies for measuring impact in processes of evidence-based policy-making.
8.	Ferguson, James. 2016. *Give a Man a Fish: Reflections on the New Politics of Distribution.* Durham, NC: Duke University Press.	Unpacking the emergence of social welfare programs that focus on direct cash payments in southern Africa, this book critically engages with the relationship between production and distribution, and untangles belonging from the neoliberal notion of making labor contributions to simply denote membership in a nation state.

9.	Gandy, Oscar. 2016. *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage*. New York, N.Y: Routledge.	Building the analytic framework of rational discrimination, this book critiques the use of probability and statistics in predictive decision-making within domains of housing, healthcare, insurance, and the criminal legal system in the United States by examining its disparate race-based outcomes.
10.	Kukutai, Tahu, and John Taylor. 2016. *Indigenous Data Sovereignty: Toward an Agenda.* Canberra: Australian National University Press.	Drawing on the United Nations Declaration on the Rights of Indigenous Peoples, this book grapples with the core challenges of articulating the meaning of data sovereignty for indigenous people and its implications for their pursuit of self-determination.
11.	Eubanks, Virginia. 2017. *Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor.* New York: St. Martin's Press.	This book weaves together critical in-depth investigation with compelling personal stories to show how the use of automated data-driven decision-making systems in delivery of public services exacerbate inequality and have deeply punitive consequences for the poor and working-class in the United States.
12.	Herd, Pamela, and Donald P. Moynihan. 2018. *Administrative Burden: Policymaking by Other Means.* New York: Russell Sage Foundation.	Tracing implementation of federal programs and controversial legislations, this book spans examples ranging from social security to voter registration laws to show that administrative burden—the cost of pursuing a public service or interacting with the government for citizens—is often a deliberate policy choice.

13.	Cheney-Lippold, John. 2019. *We Are Data: Algorithms and the Making of Our Digital Selves.* First published in paperback. New York: New York University Press.	Diving deep into the lived experience of data subjecthood, this book explores the role of algorithms in shaping our identities to ultimately argue that "who we are is what our data is made to say about us" (p. xii).
14.	Cottam, Hilary. 2019. Radical Help: *How We Can Remake the Relationships between Us and Revolutionise the Welfare State.* London: Virago Press.	Based on five practical design experiments to collaborate with communities to create spaces for sharing and listening across Britain, this book makes an argument to invest in human connection as the core resource for re-imagining and re-designing the welfare state.
15.	Gilman, Michele. 2020. "Poverty Lawgorithms: A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms on Low-Income Communities." New York: Data & Society Research Institute. https://datasociety.net/library/poverty-lawgorithms/.	Written as a guide, this report explains the ins and outs of data-centric and automated-decision making systems to poverty and civil legal services lawyers so they can better identify the source of their clients' problems and advocate on their behalf.
16.	Mattern, Shannon. 2021. *A City Is Not a Computer: Other Urban Intelligences.* Places Books. Princeton: Princeton University Press.	This book critically engages with technological visions of smart cities and computational ways of knowing the city that have increasingly come to shape urban policy and design to make a case for local, place-based ways of knowing the infrastructures and institutions that make up urban culture.

17.	Berman, Elizabeth Popp. 2022. *Thinking like an Economist: How Efficiency Replaced Equality in U.S. Public Policy.* Princeton: Princeton University Press.	The book maps the intellectual trajectory of microeconomic style of reasoning that has shaped public policy in the United States since the 1960s and how it institutionalized metrics centered on efficiency, incentives, and choice in the design of government programs.
18.	Guyan, Kevin. 2022. *Queer Data: Using Gender, Sex and Sexuality Data for Action.* London and New York: Bloomsbury Publishing.	Contending with the limitations of data practices in accounting for the everyday experiences of queer people, the book proposes tools and strategies for collecting, analyzing, and using queer data for political action.
19.	Smith, Monica L. 2022. "The Fundamentals of the State." *Annual Review of Anthropology* 51 (1): 493–508. https://doi.org/10.1146/annurev-anthro-041320-013018.	Focusing on politics, violence, literacy, and borders, this article examines the state "as a container for human interactions" and how it is constituted and expressed in ordinary configurations of power in everyday life.
20.	Pahlka, Jennifer. 2023. *Recoding America: Why Government Is Failing in the Digital Age and How We Can Do Better.* New York: Metropolitan Books ; Henry Holt and Company.	Based on hands-on experience of digitizing government services in the United States, this book describes the complex interplay of technological, organizational, and institutional challenges in translating policy into practice for service delivery and identifies strategies that center the citizen in simplifying systems and policies.
21.	Brilmyer, Gracen, and Crystal Lee. 2023. "Terms of Use: Crip Legibility in Information Systems." *First Monday* 28 (1). https://doi.org/10.5210/fm.v28i1.12935.	This special issue introduces the framework of crip legibility—"how disabled people flexibly respond to, contort, or collectively organize themselves to fit within (or be understood by) existing information systems while building new systems of resistance and care"—to contend with technoableism.

This list has been curated from our Public Zotero Library on the Datafied State. This library began as an endeavor to share resources among the participants of the Keywords on the Datafied State workshop. It has grown ever since to incorporate our ongoing literature survey, suggestions we received, and all the resources we have cited in this anthology.

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net

@datasociety

Cover and layout by Gloria Mendoza

April 2024