Confronting Data Inequality*

ANGELINA FISHER† & THOMAS STREINZ‡

Control over data conveys significant social, economic, and political power. Unequal control over data—a pervasive form of digital inequality—is a problem for economic development, human agency, and collective self-determination that needs to be addressed. This Article takes steps in this direction by analyzing the extent to which law facilitates unequal control over data and by suggesting ways in which legal interventions could lead to more equal control over data. We use the term "data inequality" to capture unequal control over data—not only in terms of having or not having data, but also in terms of having or not having the "power to datafy" (i.e., deciding what becomes or does not become data). We argue that data inequality is a function of unequal control over the infrastructures that generate, shape, process, store, transfer, and use data. Existing law often regulates

* This Article originated as a background paper for the World Development Report 2021: Data for Better Lives. We are very grateful to Victoria Adelmant, Adele Barzelay, Elettra Bietti, Nikolas Guggenberger, Niels ten Oever, Edefe Ojomo, Przemysław Pałka, David Satola, Daniel Sive, David Stein, Dimitri van den Meerssche, Christiaan van Veen, Roxana Vatanparast, and Anna Yamaoka-Enkerlin for their careful comments, incisive questions, and constructive suggestions. We also thank the participants at NYU School of Law's Information Law Institute's Privacy Research Group for their valuable feedback when we presented this project for the first time. Katie Holland, Rachel Jones, and Maxwell Votey provided indispensable editorial assistance. We are very appreciative of the care with which the editors of the *Columbia Journal of Transnational Law* have prepared this Article for publication. We draw on collaborative ideas generated at NYU School of Law's Guarini Global Law & Tech initiative, where we teach and research Global Data Law with Benedict Kingsbury: www.guariniglobal.org/global-data-law [https://perma.cc/9RPK-8BAT]. We thank our students for critical reflections and thought-provoking reactions.

† Angelina Fisher is Adjunct Professor of Law and Director for Policy and Practice of Guarini Global Law & Tech at New York University School of Law. Email: angelina.fisher@nyu.edu.

‡ Thomas Streinz is Adjunct Professor of Law and Executive Director of Guarini Global Law & Tech at New York University School of Law. Email: thomas.streinz@law.nyu.edu.

data as an object to be transferred, protected, shared, and exploited and is not always attuned to the salience of infrastructural control over data. While there are no easy solutions to the variegated causes and consequences of data inequality, we suggest that retaining flexibility to experiment with different approaches; reclaiming infrastructural control; systematically demanding enhanced transparency; pooling data and bargaining power; and developing differentiated and conditional access to data mechanisms may help in confronting data inequality more effectively going forward.

INTRODUCTION

In economic and human development narratives, the term "data inequality" typically connotes a lack of opportunities that having access to digital data might otherwise enable, such as better policy making, enhanced scientific innovation, and improved economic and social conditions. If data is a valuable resource like oil or water, how can it flow from those who have it to those who do not? This framing treats data as a valuable object to be gathered, shared, and exploited, and it tends to be concerned with law insofar as it either impedes or enables these activities.

In this Article, we suggest a different approach to data inequality. Drawing on insights from science and technology studies, media, communications, and information studies, and the emergent discipline of critical data studies, which has been pioneered by feminist and critical race scholars, we take as a starting point the position that data is not a naturally occurring phenomenon. Instead, data generation is a social practice.[1] Inequality resides not only in having or not having data, but also in having or not having the power to decide what kind of data is being generated and in what form or format, how and where it is amassed and used, by whom, for what purpose, and for whose benefit. We call this the "power to datafy."

We recognize that "data inequality" materializes in different ways, many of which are being explored within the rich scholarship on surveillance, algorithmic discrimination, digital labor, automation of the welfare state, digital mapping, and others.[2] Our focus in this Article is narrower. We employ the term "data inequality" to refer to unequal control over data, understood both in distributional terms (i.e., having or not having data) and in terms of the power to datafy (i.e., deciding what becomes or does not become data). These are by no means the only forms of digital inequality. We hope, however, that unveiling the technical, social, organizational, and legal dynamics that constitute data inequality in this more confined sense will also illuminate alternative regulatory paths for addressing the broader

1. *See* discussion *infra* Part I.

2. *See generally, e.g.*, CATHY O'NEILL, WEAPONS OF MATH DESTRUCTION (2016); SAFIYA U. NOBLE, ALGORITHMS OF OPPRESSION (2018); VIRGINIA EUBANKS, AUTOMATING INEQUALITY (2018); Rediet Abebe, Designing Algorithms for Social Good (2019) (Ph.D. dissertation, Cornell University); RUHA BENJAMIN, RACE AFTER TECHNOLOGY (2019). *See also Digital Welfare State and Human Rights Project*, CTR. FOR HUM. RTS. & GLOB. JUST., https://chrgj.org/focus-areas/technology/digital-welfare-state-and-human-rights-project [https://perma.cc/MEN7-TYER].

concerns associated with the accumulation and exercise of power through data.

In this Article, we make two key arguments. First, we posit that consideration of data inequality, as defined in this Article, requires examining the relationship between data and its constitutive infrastructures: Those who control key infrastructures necessary for data generation, transfer, and use ("data infrastructures") will be in a better position not only to accumulate data but also to determine how human life and the environment become datafied.[3] We take our inspiration for this argument from the burgeoning field of infrastructure studies, which explores the history, development, operation, maintenance, and decay of infrastructures.[4] One key insight of this interdisciplinary research agenda is to see infrastructures not merely as objects. Infrastructures are complex, relational, and highly contextual, with their effects being a function of how social, technical, and organizational elements of their assemblages relate to, intersect with, or are embedded within each other, other infrastructures, the political economy, and the law. This kind of "infrastructural analysis" can bring to light the often-less-visible enabling dynamics involved in the generation and subsequent availability of data.[5]

Second, we question the extent to which existing laws and institutions are attuned to data inequality.[6] The global "free flow" of data through the internet relies on physical infrastructures and

3. Note that our use of the term "data infrastructures" deviates from the industry usage of the term. On the distinction between its meaning in engineering and in infrastructure studies, see Florence Millerand & Karen S. Baker, *Data Infrastructures in Ecology: An Infrastructure Studies Perspective*, OXFORD RSCH. ENCYCLOPEDIA, ENV'T SCI. (Aug. 27, 2020).

4. *See* Paul N. Edwards, Geoffrey C. Bowker, Steven J. Jackson & Robin Williams, *Introduction: An Agenda for Infrastructure Studies*, 10 J. ASS'N INFO. SYS. 364, 365–66 (2009). *See generally* Benedict Kingsbury, *Infrastructure and InfraReg: On Rousing the International Law 'Wizards of Is'*, 8 CAMBRIDGE INT'L L.J. 171 (2019).

5. This approach is inspired by the InfraReg project incubated at NYU School of Law's Institute for International Law and Justice. *See InfraReg*, INST. FOR INT'L L. & JUST., http://www.iilj.org/InfraReg [https://perma.cc/6WAW-7XYF]; Kingsbury, *supra* note 4; *see also* Geoffrey C. Bowker, Karen Baker, Florence Millerand, & David Ribes, *Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment*, *in* INTERNATIONAL HANDBOOK OF INTERNET RESEARCH 97 (Jeremy Hunsinger, Lisbeth Klastrup, Matthew Allen, eds., 2010).

6. Our analysis focuses mainly on U.S. and EU law because these jurisdictions have historically shaped legal (non-) regulation of the digital economy and continue to dominate the discourse globally. In the Global Data Law project at NYU School of Law's Guarini Global Law & Tech initiative, we explore alternative pathways towards global data regulation. For more information see *Global Data Law*, N.Y.U. SCH. OF L.: GUARINI GLOBAL L. & TECH., www.guariniglobal.org/global-data-law [https://perma.cc/9RPK-8BAT].

interoperability standards. The laws facilitating and protecting this "free flow" tend to ignore considerations such as where data accumulates, between whom it flows, and who ultimately benefits.[7] In our discussion of extant laws, we note that in protecting the "free flow" of data, international economic law may entrench an unequal status quo by restricting states' abilities to localize or redistribute control over data, thereby foreclosing potential pathways towards digital development that are more attuned to data inequality. Data protection and privacy laws, as well as intellectual property laws, tend to treat data as a regulatory object, focusing predominantly on specifically delineated individual rights, but have been largely silent with respect to concentrated control over data-generating infrastructures.[8] These areas of law and the framing that they adopt for the regulation of data have not been able to address data inequality and may even entrench it. Control over data is frequently achieved through control of the relevant data infrastructures. Certain strands of antitrust and competition law and recent regulatory initiatives in the European Union seem more attuned to infrastructural control over data. They tend to incorporate, however, certain assumptions about markets, market efficiencies, and consumer welfare that may ignore broader concerns around data inequality, which ought to be addressed through other means.[9]

We focus on these areas of law—international economic law, intellectual property law, data protection and privacy law, and antitrust and competition law—because they have been most prevalently invoked as regulatory pathways for data. They are also the dominant fields from which regulatory models are being exported to or proposed for developing digital economies. Our aim is not to negate the relevance of these fields and the important contributions they can make, but rather to highlight where their framings might fall short or even undermine development objectives. There are other areas of law that are salient for data regulation (e.g., corporate and tax laws), but they are beyond the scope of this Article.[10] Overall, we seek to

7. *See, e.g.*, *infra* notes 195–197 and accompanying text.

8. *See* discussion *infra* Sections II.B & C.

9. *See* discussion *infra* Section II.D.

10. While there are numerous data-focused corporate and tax laws—such as those that mandate collection and public reporting of specific data or that require sharing of data with relevant authorities—more broadly, corporate and tax law regimes supply complex legal infrastructures that enable and support value creation and distribution by companies, including those that generate and accumulate data. *See generally* KATHARINA PISTOR, THE CODE OF CAPITAL (2019) (arguing that law constitutes the "code of capital" and is hence implicated in capitalism's inherent inequalities). *See id.* at 3 (discussing how lawyers shield assets from

overcome the siloes in which the legal regulation of data tends to be discussed, suggest alternative means of regulating data beyond extant laws, and encourage exploring alternative conceptualizations of data as a relational construct that implicates the rights and interests of different publics.

Throughout our analysis, we highlight the role of corporate power in constituting, reinforcing, and scaling data inequalities. Corporate control over data and data infrastructure is a complex phenomenon and, depending on one's normative commitments, has various positive and negative externalities. We do not suggest that corporate involvement should be avoided altogether in the interventions aimed at fostering digital economies and societies. On the contrary, we focus on corporations precisely because of the central role that they play in such interventions and because of their increasing prominence, through public-private partnerships, in the Sustainable Development Agenda[11] more broadly. Entrepreneurs, software developers, and others may derive benefits from digital infrastructures controlled by large technology companies (e.g., through services offered by cloud computing). Individuals and communities may similarly enjoy certain conveniences and pleasures afforded by data-driven technologies. At the same time, it is imperative to be fully cognizant of the effects that corporate power has on individuals, communities, and countries, particularly where such power is exercised through control over data infrastructures.[12]

Our critical analysis of concentrated corporate control over data infrastructures should not be misunderstood as an unconditional endorsement for concentrated governmental control over such infrastructures, without due regard to the particular economic, social, and political contexts in which these infrastructures are being created and deployed.[13] Nonetheless, more stringent regulation of existing data infrastructures may be necessary to ensure the development of

---

taxes); *id.* at 51–56 (analyzing the legal attributes of the modern business corporation and its potential for entity shielding). One particularly relevant example is the role that strategic incorporation plays in unequal accumulation of data and its economic value. *See* Thomas Streinz, *Data Governance in International Economic Law: Non-Territoriality of Data and Multi-Nationality of Corporations* 15 (Spring 2021) (unpublished manuscript) (on file with author).

11. G.A. Res. 70/1 (Sept. 25, 2015).

12. *See generally* Linnet Taylor & Dennis Broeders, *In the Name of Development: Power, Profit and the Datafication of the Global South*, 64 GEOFORUM 229 (2015) (exposing the growing agency of corporations as development actors).

13. We emphasize the need for context-specific interventions throughout, particularly in Part III.

data-productive, rather than data-extractive, economies. Any meaningful regulatory intervention, however, is dependent on accurate information about data infrastructures, including the provenance of and context within which data is generated, sites and mechanisms of control, and the distribution of power and interests between different actors. The notorious and somewhat paradoxical opaqueness of the most important data infrastructures can be countered through more forceful demands of transparency and data-sharing, not merely as an end in itself but as a precondition for further political and regulatory action.

To avoid the reproduction and entrenchment of data inequality, it may be necessary to reduce the dependency on data infrastructures controlled by large corporate entities, while at the same time resisting the adoption of alternative wholesale top-down data governance models that—intentionally or not—may suppress, ignore, or exploit marginalized groups. This approach may necessitate the creation, development, and support of alternative data infrastructures. Smaller, local (but also potentially transnationally aligned) actors could be empowered to make their own choices about which data to collect, and how and which data infrastructures to use and rely on.

The network effects and stark economies of scale that are hallmarks of the digital era might require the pooling of data and power to create the critical mass necessary to counter, or at least negotiate effectively with, those who possess outsized infrastructural control over data. To equalize asymmetric control over data—rather than to exacerbate it—conditional and differentiated access to data infrastructures could be developed to ensure adequate compensation and a more just recalibration of data access and benefit. In this context, the power of international organizations (and the data that they control) could be leveraged for the benefit of developing economies.

Data inequality is not a technocratic problem for which there is either a purely technical or a purely legal or even a techno-legal solution. Indeed, by undertaking an infrastructural analysis, we hope to illuminate not only the co-constitutive relationship between data law and data infrastructures but also how both are shaped by the political economy of data capitalism.[14] Institutions promoting economic and social development, as well as state actors engaged in digital policies, ought to consider the broader impacts of datafication, including those on individual welfare, development freedom, and democratic

14. In this regard, we are inspired by the works of JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L.J. 1276 (2020); PISTOR, *supra* note 10.

governance. Addressing data inequality requires a recognition of the politics of data. The context-dependency and relativity of data and data infrastructures require imagining novel ways in which publics affected by datafication can engage in public deliberation, effective contestation, and collective self-governance.

The Article develops these ideas in three Parts:

In Part I, we caution against unidimensional conceptualizations of "data" as a naturally occurring phenomenon that ought to be exploited as an economic resource. Instead, we emphasize the extent to which data is constructed through social and highly political practices. We highlight the changes to data collection, processing, transfer, and use resulting from widespread, but uneven, adoption of modern digital technologies across the globe. On this basis, we argue that unequal control over data is increasingly a function of highly concentrated corporate control over data infrastructures.

In Part II, we explore the extent to which law has facilitated unequal control over data by not addressing the infrastructural reasons for data inequality. We argue that legal interventions to address unequal control over data need to move beyond the approaches of legal data regulation that are predominantly focused on the individual protection of privacy and personal data and property-type protections of data. We acknowledge that competition law and recent regulatory initiatives in the European Union may be more attuned to the infrastructural control over data, but we also expose their inherent limitations. Embedded in our analysis is a discussion on the extent to which international economic law constrains states' abilities to localize or redistribute control over data, thereby entrenching data inequality without addressing its root causes.

In Part III, we consider possible solutions to address unequal control over data. We caution against wholesale solutions and suggest targeted interventions to confront data inequality through: (1) retaining development freedom; (2) reclaiming infrastructural control; (3) demanding enhanced transparency over data infrastructures; (4) pooling and differentiating access to data mechanisms; and (5) developing collective governance regimes of data and data infrastructures.

I. DATA INEQUALITY AS A FUNCTION OF INFRASTRUCTURAL CONTROL

We begin our analysis by challenging the unidimensional conceptualization of data as a resource, and reconceptualize data as a social practice. Interpreted from that vantage point, data inequality is more than the uneven distribution of data. The power to decide what

becomes datafied, by whom, where, how, in what form, and with what purpose is unevenly distributed. After introducing these attendant questions, we then illustrate how the power to make these determinations and the ability to accumulate data are crucially dependent on control over data infrastructures.

A. Conceptualizing Data

The term "data" is ubiquitous. Its meaning, however, differs across fields and disciplines that seek to understand and articulate what data is, what is new or different about *digital* data, and how data is transforming social, political, and economic dynamics.[15] In popular discourse, the use of metaphors is common.[16] Many analogize data to natural resources like oxygen and oil.[17]

These metaphors often conceive of data as a *natural kind*: a *resource* that exists in the wild and which can be extracted, processed, and consumed through means of industrial production. The accompanying imagery often invokes physical modalities of pipes and hoses to process and move the resource smoothly across space in order to make something visible, discoverable, traceable, observable, and ultimately calculable. These metaphors of nature operate to conjure images of data as existing *a priori* in the same way that water, air, or mineral deposits naturally exist without human intervention. This imagery is consistent with the etymology of the word data, which is

15. *See generally* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2013); ROB KITCHIN, THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES (2014); VINCENT MOSCO, TO THE CLOUD: BIG DATA IN A TURBULENT WORLD (2014).

16. For studies of metaphors of "big data," see generally Cornelius Puschmann & Jean Burgess, *Metaphors of Big Data*, 8 INT'L J. COMMC'N 1690 (2014); Jan Nolin, *Data as Oil, Infrastructure or Asset? Three Metaphors of Data as Economic Value*, 18 J. INFO., COMMC'N & ETHICS IN SOC'Y 28 (2019).

17. The Economist popularized this framing in May 2017. *See The World's Most Valuable Resource Is No Longer Oil, But Data*, ECONOMIST (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [https://perma.cc/3ACZ-34VL]. A February 2020 special report asked: "Are data more like oil or sunlight?" *See Special Report: Are Data More Like Oil or Sunlight*, ECONOMIST (Feb. 20, 2020), https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight [https://perma.cc/2SCF-C549]. Mathematician Clive Humby coined the "data is the new oil" metaphor in 2006. *See* Charles Arthur, *Tech Giants May Be Huge, But Nothing Matches Big Data*, GUARDIAN (Aug. 23, 2013, 3:21 PM), https://www.theguardian.com/technology/2013/aug/23/tech-giants-data [https://perma.cc/VD9A-YFJT].

derived from the Latin verb *dare* ("to give"): data as something that is *given*.[18] The givenness of data is thus analogized to the givenness of natural resources, which can be extracted.

Data is also often said to *flow*.[19] Technically speaking, digital data is transmitted either through light pulses or electrical signals at the behest of humans and their machines.[20] Yet, the imagery of fluidity suggests that in its state-of-nature, data moves smoothly and uninterrupted without acknowledging any human or machine agency involved in the process.[21] Data from different sources are said to aggregate into *lakes* or *pools*. When too much data accumulates so as to be unmanageable for the humans or machines extracting and processing it, the narrative of *torrents*, *floods*, and *tsunamis* of data shifts towards a need for control, as data must be cleaned, refined, and simplified to reveal its essence.[22] Once its natural force has been curbed, data is finally ready for consumption or further data creation.

Data metaphors have implications for our analysis of legal and infrastructural dimensions of data inequality. As Cornelius Puschmann and Jean Burges explain, "[T]echnological metaphors are 'never innocent' and, when deployed as part of deliberate rhetorical strategies, [they] have the potential to profoundly shape cultural and

18. Daniel Rosenberg notes that during the 18th century, the meaning of "data" shifted from something that is accepted as given to the result of experimentation, discovery, or collection. Daniel Rosenberg, *Data Before the Fact*, *in* "RAW DATA" IS AN OXYMORON 15, 36 (Lisa Geitelman ed., 2013). The change in meaning is not accidental but rather coincides with the growth and evolution of science and new modes of knowledge production that shifted away from theology to rationality, facts, evidence, and the testing of theory through experiment.

19. See, for example, our discussion of "free flows" of data in Part II.

20. *How Do Fiber-Optic Cables Transmit Data?*, SPECTRUM ENTERPRISE, https://enterprise.spectrum.com/support/faq/internet/how-do-fiber-optic-cables-transmit-data.html [https://perma.cc/9DFM-WF6Z]; *Data Transmission: What Is It?*, CDNETWORKS, https://www.cdnetworks.com/enterprise-applications-blog/everything-you-need-to-know-about-data-transmission/ [https://perma.cc/GHG8-994L].

21. The controversial idea that nonhuman artifacts or objects can have agency to be understood in relation to other nonhuman and human "actants" is usually attributed to the French sociologist and pioneer in science and technology studies, Bruno Latour. *See generally* Bruno Latour, *Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts*, *in* SHAPING TECHNOLOGY/BUILDING SOCIETY: STUDIES IN SOCIOTECHNICAL CHANGE 225 (Wiebe E. Bijker & John Law eds., 1992). *See also* Edwin Sayes, *Actor-Network Theory and Methodology: Just What Does It Mean to Say that Nonhumans Have Agency?*, 44 SOC. STUD. OF SCI. 134, 135 (2014) (clarifying the terminology).

22. On the importance and implications of classification, see generally GEOFFREY C. BOWKER & SUSAN LEIGH STAR, SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES (2000).

social practices."[23] Indeed, the metaphors outlined above have made their way into economic and political discourses.[24] The idea of data as a resource casts data as an object that can be commodified to generate further value.[25] Seeing data as a valuable resource has also meant that legal interventions have focused on questions such as: Who owns data? Who collects and processes data? How should data be shared and with whom? How should people be protected from uses and misuses of data?[26] How should concentrations of data in the hands of certain commercial actors be addressed? As we will discuss in Part II, these approaches often overlook dimensions of data inequality associated with control over data infrastructures.

Treating data as something akin to a *natural* resource has the effect of depoliticizing the processes by which data comes into existence in the first place. It not only removes human agency, but it also conceals the socio-technical practices and the surrounding politics through which phenomena are converted into a set of computationally-

23. Puschmann & Burgess, *supra* note 16, at 1697.

24. *See, e.g.*, Evîn Cheikosman, *6 Data Policy Issues Experts Are Tracking Right Now*, WORLD ECON. F. (Mar. 23, 2021), https://www.weforum.org/agenda/2021/03/6-key-issues-that-are-trending-in-data-policy-right-now [https://perma.cc/F9C8-UP99]; Nigel Cory, *The False Appeal of Data Nationalism: Why the Value of Data Comes From How It's Used, Not Where It's Stored*, INFO. TECH. & INNOVATION FOUND. (Apr. 1, 2019), https://itif.org/publications/2019/04/01/false-appeal-data-nationalism-why-value-data-comes-how-its-used-not-where [https://perma.cc/Y7SY-RUFQ].

25. The value of data can be commercial (e.g., if data itself—or the insight that it produces—is commodified and monetized), strategic, informational (e.g., for business management or policymaking), social (e.g., illuminating social problems recognizable only at scale), and so on. Data is of value, for example, to develop artificial intelligence technology reliant on large datasets, and is valued when there is a market for datasets or when the value of data is embedded in the value of a company, realized upon sale or merger. How the value of data is produced and measured remains an unsettled, but increasingly studied, question across domains such as taxation, accounting, business management, and macroeconomic analysis. *See generally* Aleksandra Bal, *(Mis)guided by the Value Creation Principle—Can New Concepts Solve Old Problems?*, 72 BULL. INT'L TAX'N 11 (2018); Chiehyeon Lim et al., *From Data to Value: A Nine-Factor Framework for Data-Based Value Creation in Information-Intensive Services*, 39 INT'L J. INFO. MGMT. 121 (2018); Wendy C.Y. Li, Nirei Makoto & Yamana Kazufumi, *Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy*, RIETI Discussion Paper Series 19-E-022, RSCH. INST. OF ECON., TRADE & INDUS. (Mar. 2019). *See also Understanding the Impact and Value of Data: Ongoing and Upcoming Projects at Open Data Watch*, OPEN DATA WATCH, https://opendatawatch.com/blog/understanding-the-value-impact-of-data [https://perma.cc/9Z3Y-V8UP] (cataloguing different studies on the "value" of data). The value of data as a corporate asset is often not being accounted for. *See* discussion *infra* Section III.C.

26. *See* discussion *infra* Part II.

manipulable measurements.[27] Remarking on data in scientific research, Sabina Leonelli observes that data "are the results of complex processes of interaction between researchers and the world, which typically happen with the help of interfaces such as observational techniques, registration and measurement devices, and the re-scaling and manipulation of objects of inquiry for the purposes of making them amenable to investigation."[28] The same observation can be made with respect to the data that is "born digital" (i.e., created immediately in binary, hence digital, code), either purposefully or incidentally.[29] The decision to capture (or measure) a particular phenomenon, process, activity or environment is unequivocally made by humans. In this way, humans design and control the means of data generation. Classifications and categorizations, formats, standards, protocols, media of storage, transport, and dissemination are all integral parts of infrastructures that make data readable, searchable, manipulatable, and transmittable via the internet. Data infrastructures and their constitutive components are thus assemblages of materialities, social norms, organizational practices, histories, ideologies, and law, in form of legal instruments, practices, and institutions. As Lauren F. Klein and Miriam Posner observe, "[D]ata sets never arrive in the world fully formed, but are assembled from tangles of historical forces and ideological motivations, as well as practical concerns."[30]

The constructed and highly contextual nature of data is well-recognized by scholars of science and technology studies ("STS"), media, communications, and information studies, and especially in the emergent discipline of critical data studies. Feminist and critical race

27. *See generally* Bruno J. Strasser & Paul N. Edwards, *Big Data Is the Answer . . . But What Is the Question?*, 32 OSIRIS 328 (2017) (deconstructing the meaning and uses of "big data" over time).

28. Sabina Leonelli, *What Counts as Scientific Data? A Relational Framework,* 82 PHIL. SCI. 810, 813 (2015). Leonelli sees data as a relational category; that is "as any product of research activities, ranging from artifacts such as photographs to symbols such as letters or numbers, which is collected, stored and disseminated in order to be used as evidence for knowledge claims." *Id.* at 817; *see also* Strasser & Edwards, *supra* note 27, at 329–30 (citing BRUNO LATOUR, PANDORA'S HOPE: ESSAYS ON THE REALITY OF SCIENCE STUDIES ch. 2 (1999), "[t]o attach the label 'data' to something is to place that thing specifically in the long chain of transformations that moves from nature to knowledge; this act of categorization marks a particular moment in time when someone thought some inscription or object could serve to ground a knowledge claim.").

29. ROB KITCHIN, *Small Data, Data Infrastructures and Data Brokers*, *in* THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES, *supra* note 15, at 31.

30. Lauren F. Klein & Miriam Posner, *Data as Media*, 3 FEMINIST MEDIA HISTORIES 1, 3 (2017).

scholars also see data not merely as a resource but as a *social practice*.[31] Viewing data through this lens has allowed these scholars to ask pertinent questions: Who gains access to and is able to extract value from data?[32] What mechanisms enable personal data to be controlled by corporations?[33] How does data production and processing shape identities, environments, and our understandings of the world?[34] These questions, in turn, have allowed for examining data not simply as a resource but also as a site of power that can reinforce (as well as subvert) existing inequalities along gender,[35] race,[36] sexuality, and class dimensions,[37] both within and across countries. A key theme that weaves through these lines of research is the importance of various practices involved in producing, accumulating, and analyzing data on democracy, freedom, self-governance, and socio-economic and political (in)equality.

These two discourses—one concerning data as a resource that can be commodified to derive value, and the other emphasizing data as a social practice—proceed largely in parallel and rarely intersect. Cognizant of this gap, we endeavor to bring these discourses into conversation with one another by adopting an infrastructural perspective. Our goal is to illustrate how data inequality is constituted through control over relevant infrastructures and to illuminate the co-constitutive relationship between infrastructures and law. These themes are explored in the ensuing Sections.

---

31. *See generally* CHRISTINE L. BORGMAN, BIG DATA, LITTLE DATA, NO DATA: SCHOLARSHIP IN THE NETWORKED WORLD xvii–xviii (2015); GEOFFREY C. BOWKER, MEMORY PRACTICES IN THE SCIENCES (2005); "RAW DATA" IS AN OXYMORON, *supra* note 18.

32. *See, e.g.*, Danah Boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO., COMMC'N & SOC'Y 662, 673 (2012).

33. *See generally, e.g.*, Julie E. Cohen, *The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy*, 31 PHIL. & TECHN. 213 (2018) (arguing that personal information is legally framed as available, valuable, and "raw"); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM (2019) (describing how Google and Facebook exercise control over personal data production).

34. *See, e.g.*, David Ribes & Steven J. Jackson, *Data Bite Man: The Work of Sustaining a Long-Term Study*, *in* "RAW DATA" IS AN OXYMORON, *supra* note 18, at 147; Noble, *supra* note 2, at 11.

35. *See generally, e.g.*, CATHERINE D'IGNAZIO & LAUREN F. KLEIN, DATA FEMINISM (2020).

36. *See generally, e.g.*, Noble, *supra* note 2.

37. *See generally, e.g.*, CHRISTIAN FUCHS & DAVID CHANDLER, DIGITAL OBJECTS, DIGITAL SUBJECTS: INTERDISCIPLINARY PERSPECTIVES ON CAPITALISM, LABOUR AND POLITICS IN THE AGE OF BIG DATA (2019).

B. Recognizing Data Inequality

In contrast to metaphors of data lakes, torrents, and firehoses, which evoke imageries of abundance, stands the metaphor of data deserts, which connote the scarcity and absence of data.[38] This imagery emphasizes the non-existence of data, spotlighting the inequality between those who have it and those who do not. The 2021 World Development Report by the World Bank recognizes this kind of data inequality.[39] The report notes that in developing economies, the unavailability of data is often due to the absence of necessary infrastructures of connectivity, storage, and processing, as well as prerequisite human labor and expertise.[40] In this Article, we seek to advance the inquiry into data inequalities[41] by foregrounding the less examined, but critically important, inequality of power to decide what data gets produced in the first place; that is, the power to decide which phenomenon gets datafied, by whom, where, and how.[42]

---

38. *See, e.g.*, Stefaan G. Verhulst & Danny Lämmerhirt, *Making Open Data More Evidence-Based: Toward a User-Centric and Interdisciplinary Research Agenda to Advance Open Data*, GOVLAB (Oct. 17, 2016), https://blog.thegovlab.org/post/making-open-data-more-evidence-based-toward-a-user-centric-and-interdisciplinary-research-agenda-to-advance-open-data [https://perma.cc/AUB3-RKW3] (noting the need to "[i]dentify and analyze 'data deserts'—where no or little data is collected and made available"); DANIEL CASTRO, CTR. FOR DATA INNOVATION, THE RISE OF DATA POVERTY IN AMERICA 2 (Sept. 10, 2014), https://www2.datainnovation.org/2014-data-poverty.pdf [https://perma.cc/B47A-MHWR] (observing that if the trend towards a data divide continues "we might even see the rise of 'data deserts'—areas of the country characterized by a lack of access to high-quality data that may be used to generate social and economic benefits").

39. WORLD BANK, WORLD DEVELOPMENT REPORT 2021: DATA FOR BETTER LIVES – CONCEPT NOTE 7 (2020), https://documents1.worldbank.org/curated/en/778921588767120094/pdf/World-Development-Report-2021-Data-for-Better-Lives-Concept-Note.pdf [https://perma.cc/R6BY-KWKM].

40. *Id.*

41. Data inequality is a relatively recent term that emerges from the expansive scholarship on digital inequalities. Jonathan Cinnamon has identified three dimensions along which data-specific inequalities (data inequalities) have emerged: (1) *access* to data, (2) *representation of the world* as data, and (3) *control* over data flows. *See* Jonathan Cinnamon, *Data Inequalities and Why They Matter for Development*, 26 INFO. TECH. FOR DEVELOPMENT 214, 215 (2020). We do not adopt his typology in this Article, although a number of issues discussed by Cinnamon will be also echoed in our analysis.

42. *See generally* Ulises A. Mejias & Nick Couldry, *Datafication*, 8(4) INTERNET POL'Y REV. (2019) (providing a brief history of the term "datafication").). Several authors examine the general idea of datafication as the conversion of complex phenomenon into data. *See generally* Katherine J. Strandburg, *Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 11 (Julia Lane et al. eds., 2014); Ira S. Rubinstein & Bilyana

Understanding the root causes for unequal data generation requires us to tease out the relationship between data and its constitutive infrastructures.

For example, so-called "data deserts" materialize when reality has not been translated (and reduced) into a computational measurement, or where proxy data (that could be used to deduce information) exists but is not generally accessible. These kinds of data gaps are neither accidental nor inevitable; they are a product of deliberate economic, social, and political choices.[43] Catherine D'Ignazio and Lisa Klein poignantly observe that "[t]he phenomenon of missing data is a regular and expected outcome in all societies characterized by unequal power relations, in which a gendered, racialized order is maintained through willful disregard, deferral of responsibility, and organized neglect for data and statistics about those minoritized bodies who do not hold power."[44]

It is often tempting to turn to mass reserves of data held by large corporate actors in attempts to irrigate data deserts. Data philanthropy, open data, and preservation of "free data flows" are examples of interventions aimed at filling the data void that is due to

---

Petkova, *Governing Privacy in the Datafied City*, 47 FORDHAM URB. L.J. 755 (2020); Joseph Savirimuthu, *Datafication as Parenthesis: Reconceptualising the Best Interests of the Child Principle in Data Protection Law*, 34 INT'L REV. L., COMPUTS. & TECH. 310 (2020).

43. *See, e.g.*, Ian Pool, *Colonialism's and Postcolonialism's Fellow Traveler: The Collection, Use and Misuse of Data on Indigenous People*, *in* INDIGENOUS DATA SOVEREIGNTY: TOWARD AN AGENDA 57, 59–68 (Tahu Kukutai & John Taylor eds., 2016) (explaining how settler colonialism displaced indigenous peoples' data infrastructures).

44. *See* D'IGNAZIO & KLEIN, *supra* note 35, at 38–39. Examples of this abound. "The Library of Missing Datasets"—a project of artist and educator Mimi Onuoha—presents a list of datasets that *ought* to exist (e.g., because they might illuminate or help address a social problem) but do not. *See* Mimi Onuoha, *On Missing Data Sets*, GITHUB (Jan. 24, 2018), https://github.com/MimiOnuoha/missing-datasets [https://perma.cc/VMX2-YTGL]. Further examples: Not one of the forty-two voluntary national reviews—state submissions for review of progress on SDGs—contained data on refugees. *See* ALICE GROSSMAN & LAUREN POST, RESCUE, MISSING PERSONS: REFUGEES LEFT OUT AND LEFT BEHIND IN THE SUSTAINABLE DEVELOPMENT GOALS (Sept. 2019), https://www.rescue.org/sites/default/files/document/4121/missingpersonreport100319.pdf [https://perma.cc/E5Z8-Y9Z4]. In the United States, immigration advocates criticized the Immigration and Customs Enforcement Agency (ICE) for not collecting data on humans contracting the SARS-CoV-2 virus while in ICE detention who might die of the illness either while detained or once released or deported. Dan Glaun, *How ICE Data Undercounts COVID-19 Victims*, PBS (Aug. 11, 2020), https://www.pbs.org/wgbh/frontline/article/how-ice-data-undercounts-covid-19-victims/ [https://perma.cc/T2V2-7PB5]. Although criminal justice and policing are increasingly data driven, there is a dearth of standardized and rigorous data about police brutality. *See* Lynne Peeples, *What the Data Say About Police Brutality and Racial Bias—And Which Reforms Might Work*, 583 NATURE, July 2020, at 22.

non-existence of data in a particular place (e.g., poverty data in Sub-Saharan Africa) or about a particular phenomenon (e.g., global ambient air pollution mortality estimates).[45] It is important to realize that whatever benefit such initiatives might hold, they also risk reproducing and accelerating inequalities of power relations that are embedded in the choices about what has become (and what was excluded from becoming) data.[46]

The decision of *what* data to produce rests fundamentally with those who control the means of data production. Data production depends on organizational practices, business models, the legal and political environment, and market pressures, among other factors. Control over the means of data production is unevenly distributed, and the interests of those in control are not necessarily aligned with societal interests. Data deserts are thus neither natural nor agentless. The power to determine what becomes datafied is related to the power to accumulate data, which is tied to the control over relevant data-generating infrastructures. Companies that enjoy control over data-gathering platforms or devices hold both the power to accumulate data and the power to determine which data is being generated through those data infrastructures. Conversely, actors that have the power to decide what data needs to be generated will often influence which data infrastructures come into existence, which, given infrastructural path dependencies, will in turn determine what will continue to be datafied (and what will not become datafied).[47]

The development of the internet and the ubiquity of devices, sensors, and platforms, coupled with increasing enhancement of computational power, have meant that choices about which data is generated are increasingly a function of infrastructures that enable data collection through devices, sensors, and platforms. Deliberate supplying of data—for example, subjects participating in research projects, individuals applying for government services, employees complying with disclosure requirements under labor and tax laws, and patients sharing their medical history to receive adequate health services—are not the only or even the primary means of data collection. Data is being generated and accumulated, at significantly greater scales, from web-browsing activities and electronic communications, as well as through the use of other internet-enabled products and services.[48] Rather than being given, data is being

45. *See infra* discussion in Parts II & III.

46. *See* Taylor & Broeders, *supra* note 12, at 232.

47. See the discussion of control over data infrastructures *infra* Section I.D.

48. *See* COHEN, *supra* note 14, at 10; LAURA DENARDIS, THE INTERNET IN EVERYTHING 4 (2020).

captured by cookies and other tracking technologies operated by companies in the data-collecting and selling business.[49] A variety of sensors in mobile phones and other personal devices like tablets and wearables also collect data about location, positioning, speed of movement, air pressure, light levels, and cellular activity levels (in addition to data about the performance of the device itself).[50] Any software application installed on these devices may collect additional types of data, with some collecting data even when the device is not being actively used.[51] Sensors are also increasingly embedded in products (e.g., cars, refrigerators, and smart TVs), physical infrastructures (e.g., bridges and water meters), and even in biological matter (e.g., biosensors).[52] Rather than being periodic, planned, and purpose-focused, the collection of data is increasingly continuous and ubiquitously deployed for a multitude of uses and reuses.[53]

Many of the relevant decisions *about what becomes data* (i.e., what is being datafied) are made by commercial actors who exercise varying degrees of control over data infrastructures through which data is ultimately generated and processed. In practice, data is generated at a significantly greater scale and speed through platforms and devices than through surveys or scientific samplings.[54] As a result, the various inequalities and power imbalances that exist within a given political economy are reproduced through data on a very large scale. Actors

---

49. *See* WOLFIE CHRISTL, CORPORATE SURVEILLANCE IN EVERYDAY LIFE: HOW COMPANIES COLLECT, COMBINE, ANALYZE, TRADE, AND USE PERSONAL DATA ON BILLIONS 73–75 (June 2017), https://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf [https://perma.cc/BN46-GJXG].

50. David Nield, *All the Sensors in Your Smartphone, And How They Work*, GIZMODO (June 29, 2020, 10:38 AM), https://gizmodo.com/all-the-sensors-in-your-smartphone-and-how-they-work-1797121002 [https://perma.cc/4C8L-27FL].

51. Jason Cohen, *These Apps Collect the Most Personal Data*, PCMAG (Jan. 11, 2022), https://www.pcmag.com/news/sick-of-data-collection-try-these-apps-instead [https://perma.cc/H9N6-7LEE].

52. *See generally* Patrika Mehrotra, *Biosensors and Their Applications – A Review*, 6 J. ORAL BIO. & CRANIOFACIAL RSCH. 153 (2016).

53. The relativity and variability of data poses challenges for data regulation by law, as we explore below with regard to "personal data" (see discussion *infra* Section II.C) and complications for competition law analysis arising from cross-sectoral data use (see discussion *infra* Section II.D).

54. Increasingly, so-called "digital trace" data is also being combined with survey data. *See generally* Sebastian Stier, Johannes Breuer, Pascal Siegers & Kjerstin Thorson, *Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field*, 38 SOC. SCI. COMPUT. REV. 503 (2020). *See generally* MATTHEW SALGANIK, BIT BY BIT: SOCIAL RESEARCH IN THE DIGITAL AGE (2017) (exploring the potential of "big data" for social science).

who control data infrastructures are often in a favored position to accumulate data. As data can be reused, including for purposes other than the original, data reserves become attractive gap-fillers for users who lack their own infrastructures for data collection. For instance, data philanthropy movements encourage data "donations" as a way of remedying situations of data deserts.[55] Calls for open data similarly aim to increase the availability of data to wider constituencies.[56] Without rendering normative judgments on the success of these initiatives,[57] we note here that the proliferation of data produced under concentrated control over its means of production puts those who possess such control in a privileged position in determining how the world is to be represented, (re)shaped, and governed.[58]

---

55. *See* Yafit Lev Aretz, *Data Philanthropy*, 70 HASTING L.J. 1491, 1493 (2019).

56. *See*, *e.g.*, STEFAAN G. VERHULST & ANDREW YOUNG, OPEN DATA DEMAND: TOWARD AN OPEN DATA DEMAND ASSESSMENT AND SEGMENTATION METHODOLOGY 5, https://thegovlab.org/static/files/publications/Data+Demand.pdf [https://perma.cc/35TT-VPXD]; STEFAAN G. VERHULST & ANDREW YOUNG, OPEN DATA IN DEVELOPING ECONOMIES: TOWARD BUILDING AN EVIDENCE BASE ON WHAT WORKS AND HOW pt. 3 (2017), https://odimpact.org/files/odimpact-developing-economies.pdf [https://perma.cc/T3XE-4ZU3] (discussing the impact of open data when used by different constituencies). More generally, see the Open Government Data project of the Organization for Economic Co-operation and Development (OECD). *Open Government Data*, OECD, https://www.oecd.org/gov/digital-government/open-government-data.htm [https://perma.cc/NUY7-YWBD].

57. *See* discussion *infra* Section II.B (discussing who benefits from making public data available as open data) & Section III.D (contrasting differential access to data solutions with open data).

58. The politics of knowledge production in a digitalized world is, of course, not a new topic. Critical information studies interrogate "structures, functions, habits, norms, and practices that guide global flows of information and cultural elements." *See* Siva Vaidhyanathan, *Afterword: Critical Information Studies: A Bibliographic Manifesto*, 20 CULTURAL STUD. 292, 292 (2006). Critical data studies track "the ways in which data are generated, curated, and how they permeate and exert power on all manner of forms of life." *See* Andrew Iliadis & Federica Russo, *Critical Data Studies: An Introduction*, BIG DATA & SOC'Y, Oct. 2016, at 2. We do not reproduce these discussions here since no deservingly full account would be possible. *See also* GOVERNANCE BY INDICATORS ch. 1 (Kevin E. Davis, Angelina Fisher, Benedict Kingsbury & Sally Engle Merry eds., 2012) (examining ways in which quantitative rank-able indicators act as technologies of governance by defining problems, theories of change, and proposed solutions, thereby influencing ways in which political, social, and economic decisions are being made). These critiques of knowledge production are distinct from a critique of "big data" on the grounds that it lacks proper sampling and is thus inherently biased. *Contra* Nobuo Yoshida, *Revolutionizing Data Collection: From "Big Data" to "All Data,"* WORLD BANK BLOGS (Dec. 11, 2014), https://blogs.worldbank.org/developmenttalk/revolutionizing-data-collection-big-data-all-data [https://perma.cc/Y97E-VG47].

Those who possess data also often control the terms under which others may (or may not) access and use it. Thus, apart from nonexistence in the first place, data that is available, in a sense that it *has been collected by someone*, can nonetheless be inaccessible or accessible only to certain constituencies and/or only under certain circumstances and conditions. Accessibility can be regulated by law and technical means, or it can be a function of organizational dynamics.

Governments can mandate access to data, for example, through statutory and regulatory mechanisms requiring public reporting or the provision of data to specific government agencies, competitors, or consumers.[59] Commercial entities can employ legal instruments like contracts or licenses, often rooted in regimes of personal and intellectual property, to preclude access to data or to establish conditions for such access.[60] Even without recourse to "legal technologies,"[61] access to data can be regulated through technical means. For example, companies can use cryptography within their hardware and software to minimize unauthorized access to stored data.[62] They can also implement encryption protocols (e.g., hypertext transfer protocol secure or "https") to protect data flows from unauthorized access. Using these tools, companies can, for example, prevent access to or devalue information that may be intercepted by

59. Examples of these abound in different contexts at international, regional, national and local levels. *See*, *e.g.*, International Maritime Organization [IMO], *Maritime Env't Protection Comm. Res. 278(70), Amendments to the Annex of the Protocol of 1997 to Amend the International Convention for the Prevention of Pollution from Ships, 1973, as Modified by the Protocol of 1978 Relating Thereto: Amendments to MARPOL Annex VI*, Regulation 22A (Oct. 28, 2016) (requiring ships of 5,000 gross tonnage or above to start collecting and reporting data to an International Maritime Organization database from 2019); Regulation 2019/1150 of the European Parliament and of the Council of 20 June 2019 on Promoting Fairness and Transparency for Business Users of Online Intermediation Services, 2019 O.J. (L 186); Regulation No 715/2007 of the European Parliament and of the Council of 20 June 2007, 2007 O.J. (L 171), art. 6 (mandating access to vehicle repair and maintenance information); Digital Economy Act 2017-30 (Gr. Brit.) (mandating data sharing across public sector); New York City Council, Data on Orders Placed Through Third-Party Delivery Services, Int. No. 2311-A (Aug. 29, 2021).

60. *See* discussion *infra* Section II.B.

61. This terminology highlights the malleability of law and the role of lawyers in using and adapting legal instruments to further the interests that they represent. *See* Kevin E. Davis, *Contracts as Technology*, 88 N.Y.U. L. REV. 83, 85 (2013) (analyzing innovation in contractual documents).

62. JAMES GRIMMELMANN, INTERNET LAW: CASES AND PROBLEMS 39–43 (11th ed. 2021).

the government.[63] Similarly, however, they can, if compelled by law or otherwise, provide access to users' data to a third party without users' knowledge by adding "backdoors" into software, particularly in cloud-provided services where the users are not offered a choice of whether to accept or download an update.[64]

Access to data can also be regulated through the adoption of particular standards. Making data available only in certain formats can make data practically inaccessible for certain uses or users.[65] For example, data that is not machine readable is harder to aggregate with other data and cannot be used as easily for machine-learning purposes. As Tarleton Gillespie notes, "A technology that facilitates some uses can do so only by inhibiting others."[66] Thus, for example, choice of file formats and coding protocols inherently determines not only which data is produced, but also who gets to participate in its production and use.[67] Similarly, platforms' boundary resources (i.e., resources which facilitate the use of core platform functionality to build applications)—such as software development kits ("SDKs") and application programming interfaces ("APIs")—control who can access the platforms' hardware, operating systems, and data (i.e., metrics,

63. For example, Google and Microsoft started deleting many identifiers associated with web searches from their databases after six to nine months to provide some level of anonymity to users, simultaneously diminishing the risks associated with unauthorized surveillance and data breaches. *See* Peter Fleischer et al., *Another Step to Protect User Privacy*, GOOGLE BLOG (Sept. 8, 2008), https://googleblog.blogspot.com/2008/09/another-step-to-protect-user-privacy.html [https://perma.cc/5U2U-9DRQ]; Kevin J. O'Brien, *Microsoft Puts a Time Limit on Bing Data*, N.Y. TIMES (Jan. 19, 2020), https://www.nytimes.com/2010/01/20/technology/companies/20search.html?mtrref= [https://perma.cc/4XJF-7XAH].

64. Christopher Soghoian, *Caught in the Cloud: Privacy, Encryption, and Government Back Doors in the Web 2.0 Era*, 8 J. TELECOMM. & HIGH TECH. L. 359, 419–20 (2010).

65. In the United Kingdom, nearly 16,000 Sars-CoV-2 infections went initially unreported, complicating contact tracing efforts, because Public Health England, the national health agency, had used an outdated file format to collate test results. *See Covid: Test Error 'Should Never Have Happened' – Hancock*, BBC (Oct. 5, 2020), https://www.bbc.com/news/uk-54422505 [https://perma.cc/7G6T-QZ33].

66. TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 179 (2018).

67. For instance, content producers who work with video formats different from those required by YouTube, or who produce videos at length exceeding that allowed by YouTube, may be unable to either post or access content and may be forced to alter or subtly shape the videos themselves. *Id.*

analytics, and metadata) and under what conditions.[68] Such access is critical to creating complementary applications or services.[69]

Protocols can also be deployed to prevent third parties from harnessing users' data. For example, in March 2020, Apple released a new version of its Safari browser which blocked all third-party cookies by default, thereby preventing other companies from tracking users across multiple websites.[70] Apple also allowed users to see which trackers had been blocked.[71] According to one report, the Safari browser blocked ninety trackers in five minutes, a vast majority of them being Google analytics.[72] Google has also announced that it will move towards third-party cookie blocking, but over a period of two years.[73] These companies regulate, via control over infrastructure, who is able to accumulate and derive value from data about internet users as well as when and how they are able to access it.[74]

This kind of infrastructural control over data flows is not dependent on legal technologies, though violations of contractually agreed policies are sometimes cited to justify these moves. For example, in 2019, Apple used its control over the AppStore and the operating systems of iPhones to block Facebook from operating a data collection app, alleging that it violated the terms of its enterprise

68. Ahmad Ghazawneh & Ola Henfridsson, *Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model*, 23 INFO. SYS. J. 173, 174 (2013).

69. *Id*.

70. Nick Statt, *Apple Updates Safari's Anti-Tracking Tech with Full Third-Party Cookie Blocking*, THE VERGE (Mar. 24, 2020, 3:07 PM), https://www.theverge.com/2020/3/24/21192830/apple-safari-intelligent-tracking-privacy-full-third-party-cookie-blocking [https://perma.cc/2JLR-H3TX].

71. *Safari User Guide: See Who Was Blocked from Tracking You in Safari on Mac*, APPLE, https://support.apple.com/guide/safari/see-who-tried-to-track-you-ibrw35004465/mac [https://perma.cc/B5UB-8K8V].

72. John Koetsier, *Apple's New Browser Blocked 90 Web Trackers in 5 Minutes,* FORBES (Sept. 17, 2020, 2:04 PM), https://www.forbes.com/sites/johnkoetsier/2020/09/17/apples-new-browser-blocked-90-web-trackers-in-5-minutes/?sh=1a239c1b11eb [https://perma.cc/R8PL-6VKM].

73. Dieter Bohn, *Google Delays Blocking Third-Party Cookies in Chrome until 2023*, THE VERGE (June 24, 2021, 9:21 AM), https://www.theverge.com/2021/6/24/22547339/google-chrome-cookiepocalypse-delayed-2023 [https://perma.cc/AJ32-QZW3].

74. For a summary of the debates, see Dieter Bohn, *Google to 'Phase Out' Third-Party Cookies in Chrome, But Not for Two Years*, THE VERGE (Jan. 14, 2020, 11:00 AM), https://www.theverge.com/2020/1/14/21064698/google-third-party-cookies-chrome-two-years-privacy-safari-firefox [https://perma.cc/R5WG-4SVD].

certificate policy.[75] In the summer of 2020, Apple updated its operating system to limit the data-gathering ability of third parties, making it more difficult for advertisers and platforms dependent on advertising revenue, such as Facebook, to track Apple customers, citing privacy concerns while benefitting from continued preferential access to user data.[76] In another example, Google used its control over the desktop and mobile versions of its Chrome browser to nudge users towards encrypted domain name system (DNS) services,[77] on which the resolution of domain names (e.g., nyu.edu) into numbers (i.e., internet protocol addresses, such as 216.165.47.10) depends. While supported by many privacy and security advocates, this move deprived Internet Service Providers (ISPs), which had traditionally provided unencrypted DNS resolution services, of the ability to monitor the internet use of their customers and also imperiled regulatory interventions dependent on such ISP-enabled monitoring.[78] All these

75. Josh Constine, *Apple Bans Facebook's Research App that Paid Users for Data*, TECHCRUNCH (Jan. 30, 2019, 10:44 AM), https://techcrunch.com/2019/01/30/apple-bans-facebook-vpn/#:~:text=The%20Research%20app%20asked%20users,going%20through%20the%20App%20Store [https://perma.cc/W6DH-Q692].

76. Reed Albergotti & Elizabeth Dwoskin, *Apple Makes a Privacy Change, and Facebook and Advertising Companies Cry Foul*, WASH. POST (Aug. 28, 2020, 11:59 AM), https://www.washingtonpost.com/technology/2020/08/28/facebook-apple-ios14/ [https://perma.cc/JZP6-3T4C]; D. Daniel Sokol & Feng Zhu, *Harming Competition and Consumers Under the Guise of Protecting Privacy: An Analysis of Apple's iOS 14 Policy Updates* 1 (U.S.C. L. Legal Stud. Rsch. Paper, No. 21-27, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3852744# [https://perma.cc/YL2C-7BKU] (arguing that Apple's operating system update represents an anti-competitive strategy disguised as a privacy-protecting measure).

77. Michael Grothaus, *Google Chrome Gets Major Privacy Boost: Here's How to Enable DNS-over-HTTPS*, FAST CO. (May 20, 2020), https://www.fastcompany.com/90507324/google-chrome-gets-major-privacy-boost-heres-how-to-enable-dns-over-https [https://perma.cc/F6E4-GTPH]; Arooj Ahmed, *Google Is Extending Secure DNS to Chrome 85 for Android Soon*, DIGIT. INFO. WORLD (Sept. 4, 2020, 4:22 PM), https://www.digitalinformationworld.com/2020/09/google-is-extending-secure-dns-to-chrome-85-for-android-soon.html [https://perma.cc/RK3A-JNZ4]. Mozilla's Firefox browser made a similar move, but the lobbying campaign by ISPs focused on Google. *See* Jon Brodkin, *Firefox Turns Encrypted DNS on by Default to Thwart Snooping ISPs*, ARS TECHNICA (Feb. 25, 2020, 6:00 AM), https://arstechnica.com/information-technology/2020/02/firefox-turns-encrypted-dns-on-by-default-to-thwart-snooping-isps [https://perma.cc/WF2V-B59L].

78. Mark Jackson, *Google, UK ISPs and Gov Battle Over Encrypted DNS and Censorship*, ISPREVIEW (Apr. 22, 2019, 9:49 AM), https://www.ispreview.co.uk/index.php/2019/04/google-uk-isps-and-gov-battle-over-encrypted-dns-and-censorship.html [https://perma.cc/K8E3-A7VQ]; *DNS Shakeup Could Kill ISP Filters*, IT PRO (June 22, 2019), https://www.itpro.co.uk/network-internet/33917/dns-shakeup-could-kill-isp-filters [https://perma.cc/89Y4-S8X4].

developments illustrate how corporate infrastructural control over data flows complicate the "free flow of data" narrative.[79]

Even where data exists and is not made inaccessible through legal or technical means, usability of data can be reduced when organizations fail to maintain or update relevant infrastructures or, through path dependencies, insist on retaining legacy systems for too long. During the COVID-19 pandemic in spring 2020, U.S. Social Security authorities struggled to implement increased unemployment benefits because their mainframe computers ran a sixty-year-old programming language called COBOL for which programmers were lacking.[80]

These examples illustrate how control over data infrastructures conveys power over datafication and the acquiring and amassing of data. In the ensuing Section, we explore infrastructural control in more detail.

C. Disentangling Infrastructures

"Infrastructure" is a notoriously ill-defined and ubiquitous term—not unlike data. There is and can be no uniform definition of infrastructure, as infrastructures are inherently context-dependent and relational: what is infrastructural for some might not be infrastructural for others. As mentioned at the outset, we draw on infrastructure studies, which conceptualize infrastructures not merely as objects.[81] For instance, fiber-optic cables or data centers are not infrastructures in and of themselves. They become *infrastructural* only when considering the social, economic, and political contexts within which processes, practices, norms, and rules bring about their existence, geographical positions, ownership and control structures, and operations, thereby establishing connections to other components and infrastructures.[82] The upshot of this approach for legal scholars is that all these dimensions are not just entangled with each other—so that their disentanglement might provide analytical value—but also with

79. *See* discussion *infra* Section II.A.

80. Ian King, *An Ancient Computer Language Is Slowing America's Giant Stimulus*, BLOOMBERG (Apr. 13, 2020, 5:00 AM), https://www.bloomberg.com/news/articles/2020-04-13/an-ancient-computer-language-is-slowing-america-s-giant-stimulus [https://perma.cc/H8W9-R2T9].

81. For additional background on infrastructure studies, see *supra* note 4 and accompanying text.

82. On using infrastructure studies for legal analysis, see *supra* note 5 and accompanying text.

various legal instruments, whether public or private, international or domestic.[83]

Our reference to *data infrastructures* in this Article calls attention to the technical, social, political, organizational, and legal dimensions of complex assemblages that capture, generate, categorize, standardize, aggregate, modify, (re)assemble, (re)interpret, transfer, or use data for a variety of purposes. The technical dimension of such data infrastructures consists of various components of digital infrastructure on different layers (e.g., hardware and software, localized or decentralized computing and storage facilities, internetworking and cloud computing capabilities). The social dimension underscores the variegated human (and human-machine) interactions implicated in such data-related activities, such as social practices, community norms, or the individual behavior of data subjects, who intentionally or inadvertently form part of data infrastructures. The organizational dimension, which we emphasize by focusing on corporate actors, interrogates which structures and processes hold data infrastructures together and explores how governance and decision-making structures shape decisions about datafication. The legal dimension calls attention to ways in which legal regimes and technologies help constitute and shape infrastructure and are, in turn, shaped by it. The political dimension illuminates not just the politics of infrastructures, but how infrastructures constitute and implicate different publics.[84]

Data infrastructures exist in different contexts and on varied scales. They entail different types of data and implicate a diversity of actors and interests. Some data infrastructures are domain specific (e.g., health information exchanges), whereas others are more generic (e.g., behavioral user data). Some are accessible by all (e.g., the Humanitarian Data Exchange),[85] while others are closed and made available only to a circumscribed group of people (e.g., student performance records available only to parents and teachers). Some data infrastructures have a for-profit use purpose (e.g., Facebook's behavioral data infrastructure for targeted ads),[86] others can be used

83. The infrastructural analysis here is hence the basis for our discussion *infra* in Part II.

84. *See* Benedict Kingsbury & Nahuel Maisley, *Infrastructures and Laws: Publics and Publicness*, 17 ANN. REV. L. & SOC. SCIS. 353, 364–66 (2021) (discussing divergences between legal and infrastructural publics).

85. THE HUMANITARIAN DATA EXCHANGE, https://data.humdata.org/ [https://perma.cc/FG6Q-PRYH].

86. Kurt Wagner, *This Is How Facebook Uses your Data for Ad Targeting*, VOX (Apr. 11, 2018, 6:00 AM), https://www.vox.com/2018/4/11/17177842/facebook-advertising-ads-explained-mark-zuckerberg [https://perma.cc/4GR7-7TWM].

for any purpose (e.g., open government data infrastructures),[87] and yet others are made available on the promise of non-commercial use (e.g., Equinor's data on the decommissioned Volve oil field in the North Sea).[88] Some data infrastructures are managed by government entities (e.g., population census), while others are managed by private commercial actors (e.g., data.world—a public benefit corporation that, among other offerings, hosts a large collection of open data)[89] or non-profit bodies (e.g., Investigative Reporting Workshop, an independent, nonprofit newsroom, which hosts large public datasets as part of its Accountability Project).[90] Many data infrastructures involve both commercial and not-for-profit actors, thereby blurring the "public" and "private" distinction.[91] Some data infrastructures are intensely local (e.g., the DECODE project, which was piloted in Amsterdam and Barcelona),[92] while others are transnational (e.g., the United Nations Statistical Division's Federated Data Model for data related to the Sustainable Development Goals).[93] Often, however, data infrastructures are *both* local (e.g., collecting data about individuals or communities) and transnational (e.g., data collected and processed by multinational corporations, stored across data centers in different jurisdictions, or used by constituencies dispersed around the globe). Many data infrastructures generate data purposefully, sometimes at regular or specified intervals, with the entire process of collection,

87. The repurposing of governmental data is a key narrative in the open data movement. *See* JONATHAN GRAY, TOWARDS A GENEALOGY OF OPEN DATA 9 (Sept. 3, 2014), https://dx.doi.org/10.2139/ssrn.2605828 [https://perma.cc/MH8J-VLM9]; *see also* discussion *infra* Section III.D.

88. *Volve*, EQUINOR, https://www.equinor.com/en/what-we-do/norwegian-continental-shelf-platforms/volve.html [https://perma.cc/P89P-4KW3].

89. Data.world provides a data catalogue of open datasets. DATA.WORLD, https://data.world/datasets/open-data [https://perma.cc/4SY7-8RGB].

90. INVESTIGATIVE REPORTING WORKSHOP, https://investigativereportingworkshop.org/ [https://perma.cc/Q8ET-7UC2].

91. Consider, for example, Data Partnership—an initiative of the World Bank (an inter-governmental organization) in partnership with private actors, including public benefit corporations like AtlasAI—that aims to leverage data to further development aims. DATA PARTNERSHIP, https://datapartnership.org/about/ [https://perma.cc/BU6J-RK9R]. More generally, on the blurring and politics of the public-private distinction, see the seminal work of Duncan Kennedy, *The Stages of the Decline of the Public/Private Distinction,* 130 U. PA. L. REV. 1349 (1982).

92. DECODE is a Europe-wide consortium. *See* DECODE, https://decodeproject.eu/have-more-questions.html [https://perma.cc/P2TD-5FN6].

93. UNITED NATIONS STAT. DIV., THE FEDERATED INFORMATION SYSTEM FOR THE SDGS: FROM VISION TO SCALE 1–2 (2019), https://unstats.un.org/unsd/statcom/50th-session/side-events/20190307-1L-Federated-Information-System-for-the-SDGs.pdf [https://perma.cc/3EE9-6AUE] [hereinafter FEDERATED INFORMATION SYSTEM].

aggregation, processing, and use governed by rules, norms, and laws (e.g., electronic health records). Increasingly, however, data is also being generated passively, continuously, and incidentally through platforms, sensors, and devices.[94]

The new means, scope, and speed of data generation have resulted in the creation of sprawling, jurisdictionally unbounded data infrastructures, connecting sensors and other devices, extracting, cleaning, storing, aggregating, and otherwise processing data that implicates individuals, communities, and environments. Different entities may exercise control at different points of the process by which a set of measurements about physical phenomena become usable and analyzable data. Thus, different entities may have different types of control. For example, producers of hardware that collects measurements have control over *how* those measurements are collected. When that hardware is sold, leased, or provided as a service to another party pursuant to a contractual arrangement, the recipient from that point on will exercise control, for instance, over where the hardware is located. A company that designs software used to aggregate, clean, and otherwise process data exercises control over data, including through standards and formats. Similarly, that software can be either sold or provided as a service, thereby transferring certain elements of control over data to their recipient. Finally, companies that operate data management platforms will exercise control over how processed data is analyzed, how insights of the analysis are presented, what type of access is given to the users, for what purposes, and so on.

These observations require us to confront a paradox: how is it even possible to exercise control over large-scale, highly complex, and often distributed data infrastructures?

D. Identifying Control Over Infrastructure

While control over infrastructures is rarely absolute, control over critical elements (e.g., a particular protocol, application, or

94. Ambient data collection challenges a number of established binaries, including natural/digital (e.g., think digital twins), material/virtual (e.g., think 3-D printing, augmented and virtual reality devices), and bodies/objects (e.g., think wearables and implanted devices). These are human perceptions of cyber-embeddedness that can replicate or fill-in-the-gaps in perceptions of material reality (examples include wearables, biometric ID, digital monitoring, implanted devices). On the relationship between humans and machines more generally, see the seminal work of DONNA HARRAWAY, *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*, *in* SIMIANS, CYBORGS AND WOMEN: THE REINVENTION OF NATURE 149–81 (1991).

operating system) can be sufficient to secure and leverage control over the power to datafy and the resulting data.

The internet—a technically distributed but increasingly economically-centralized data infrastructure with centralized points of governance and control—has been (and continues to be) foundational, not only for the development and growth of platforms but also in enabling control over ever-more-expansive data infrastructures. The internet enables data transfers and serves as a foundational infrastructure for cloud computing which provides the overall organization for many data infrastructures today.[95] Cloud computing enables the storage and processing of very large datasets in distributed fashion (leading to cost-efficient locating of data) and offers enterprise services that open up new possibilities for data use.[96] Software (and specifically software that extracts, cleanses, aggregates, processes, and analyzes data) can modify and extract data to convert it into economic value. This trifecta has not only facilitated and accelerated the extraction of value from data; it has also opened possibilities in the development of artificial intelligence, with wide-ranging applications, which, in turn, has fueled demand for ever larger amounts of data, while often ignoring negative externalities.[97]

Platforms are often in a favorable position to generate and accumulate data.[98] The cross-over between platform studies and

95. *Cf.* NICK COULDRY & ULISES A. MEJIAS, THE COSTS OF CONNECTION: HOW DATA IS COLONIZING HUMAN LIFE AND APPROPRIATING IT FOR CAPITALISM 39 (2019) ("[T]he Cloud Empire is the *what*, the overall organization of resources and imagination that emerges from the practices of data colonialism.").

96. On cloud computing, see generally REGULATING THE CLOUD: POLICY FOR COMPUTING INFRASTRUCTURE (Christopher S. Yoo & Jean François Blanchette eds., 2015).

97. Former Google CEO Eric Schmidt famously observed: "[B]ig data is so powerful, nation states will fight [over it]." Rob Price, *Alphabet's Eric Schmidt: 'Big Data Is So Powerful, Nation States Will Fight' Over It*, BUS. INSIDER (Mar. 9, 2017, 1:51 PM), https://www.businessinsider.com/google-eric-schmidt-countries-will-fight-over-big-data-alphabet-cloud-2017-3?r=DE&IR=T [https://perma.cc/2R8Z-VVZW]. On the downsides of ever larger datasets for natural language processing, *see generally* Emily M. Bender et al., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, FACCT '21 (Mar. 3–10, 2021), https://dl.acm.org/doi/pdf/10.1145/3442188.3445922 [https://perma.cc/Q7T8-P868].

98. The term digital "platform" has been used to refer to a variety of different online structures—and corresponding business models—that enable a wide range of activities between different actors for different purposes. For a discursive take on the meaning of "platform," see generally Tarleton Gillespie, *The Politics of 'Platforms*', 12 NEW MEDIA & SOC'Y 347 (2010). In communication studies, platforms denote "sites and services that host, organize, and circulate users' shared content or social exchanges for them; without having produced or commissioned [the majority of] that content; beneath that circulation, an

infrastructure studies illuminates how platforms have acquired infrastructural significance in a range of economic and social functions and have become not only dominant data-holders but also significant shapers of the digital world.[99] By serving as intermediaries and by bundling up various services, platform companies may cement and consolidate infrastructural control over data by creating "walled gardens" within which data is generated, accumulated, concentrated, and protected.[100] Additionally, their corporate organization gives them legal grounding for the "platformization of infrastructures" and the "infrastructuralization of platforms."[101]

In other words, platform companies—themselves infrastructures for e-commerce, communication, software services, or other transactions—are also data infrastructures.[102] The congruence between corporate control over dominant platforms and the resultant data-generating and -shaping capacity leads to concentrated control over data and results in an outsized power to datafy. This is not to say that only platform companies enjoy infrastructural control over data. There are other corporate actors that are not platforms—understood as intermediaries for two-sided markets—which control important and large-scale data infrastructures. Examples of non-platform corporate actors include network operators and networking service providers (e.g., Cloudflare), networking equipment vendors (e.g., Nokia, Huawei, Ericsson, and Juniper), or manufacturers and operators of internet-of-things (IoT) devices (e.g., Honeywell and Cisco). Yet, platforms companies are particularly potent data generators, as the following examples illustrate.

---

infrastructure for processing that data (content, traces, patterns of social relations) for customer service and for profit." *See* Tarleton Gillespie, *Regulation of and by Platforms*, *in* THE SAGE HANDBOOK OF SOCIAL MEDIA 254, 254 (Jean Burgess, Thomas Poell & Alice Marwick eds., 2018). In competition law and economics, platforms are often analyzed as 'two-sided markets' "when (1) a single transaction takes place between two different groups of users connected by the platform, and (2) the numerosity of each group of users creates reciprocal inter-side positive externalities." Giacomo Luchetta, *Is the Google Platform a Two-Sided Market?*, 10 J. COMPETITION L. & ECON. 185, 192 (2013). According to the OECD, a platform facilitates interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the internet. *See* OECD, AN INTRODUCTION TO ONLINE PLATFORMS AND THEIR ROLE IN THE DIGITAL TRANSFORMATION 11 (2019), https://doi.org/10.1787/53e5f593-en [https://perma.cc/2MHV-AWGB].

99. *See generally* Jean-Christophe Plantin et al., *Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook*, 20 NEW MEDIA & SOC'Y 293 (2018).

100. *Id.* at 301–06.

101. *Id.* at 306–07.

102. Note that this is sometimes conveyed by saying that X is not just a Y company but also a "data company".

E-commerce platforms (e.g., Amazon, Shopify, Alibaba, and JD.com), social media platforms (e.g., Facebook, YouTube, Twitter, and WeChat), search engines (e.g., Google and Baidu), and smartphones' operating systems and app marketplaces (e.g., Apple's iOS with the App Store or Google's Android with Google Play) generate gigantic amounts of data, which their corporate owners hoard and use to shape the world.[103] These companies are not only infrastructural for all forms of connectivity (technical, economic, social, and political), but they are also data infrastructures that create identities, foster social practices, shape economic relations, and engender new industries. These companies' abilities to produce, collect, analyze, and intensely protect "their" data from competitors is supported not solely, or even mainly, by resorting to legal protections. They can exercise control over different layers or components of data infrastructures and acquire further data infrastructures as needed.

Consider, for example, e-commerce platform companies that amass behavioral data by surveilling the commercial activity of sellers and buyers. These e-commerce platforms have invested heavily in logistics, payment, and even microfinance infrastructures, thereby obtaining control over key infrastructures for the online trade of goods and services.[104] Amazon's platform operates as an infrastructure for commerce,[105] which allows it to capture data of buyers and sellers. The company uses clickstreams (i.e., digital breadcrumb trails) to monitor which sites users come from, how they move through Amazon's pages, and where they go to next.[106] Amazon and its peers,

---

103. Data brokers, an emerging and highly profitable industry, feed off other platforms' data generating capacity as they scrape websites, buy data captured and/or aggregated by others, reaggregate and repackage them and offer this data or "data insights" derived from it for sale. *See* Leanne Roderick, *Discipline and Power in the Digital Age: The Case of the US Consumer Data Broker Industry*, 40 CRITICAL SOCIO. 729, 729 (2014).

104. For an interesting case study of Alibaba.com, see generally SEUNG HO PARK & ZIQIAN ZHAO, ALIBABA GROUP: FOSTERING AN E-COMMERCE ECOSYSTEM (2016). In 2011, Alibaba announced that it was spending close to $4.5 billion USD on logistics and on building out integrated warehouse networks across China. Alibaba started to offer loan applications to retailers online, using Alibaba-developed credit assessment models as well as behavioral data generated by the sellers' daily transactions. The resulting credit and risk assessments produced additional data. *See Alibaba Establishes Small Loans Lender in Chongqing*, CHINA BUS. NEWS (June 2, 2011), http://bi.gale.com.ezproxy.cul.columbia.edu/essentials/article/GALE%7CA259621401?u=nysl_oweb [https://perma.cc/D79C-8XRJ].

105. K. Sabeel Rahman, *The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1673 (2018).

106. *See* Leo Kelion, *Why Amazon Knows So Much About You*, BBC, https://www.bbc.co.uk/news/extra/CLQYZENMBI/amazon-data [https://perma.cc/P3K7-9AQC]; Lina M. Khan, *Sources of Tech Platform Power*, 2 GEO. L. TECH. REV. 325, 329–31

like the Chinese giant Alibaba, have developed corporate organizations to integrate their operations across markets, which have allowed these platforms to establish advantageous positions in adjacent markets,[107] which in turn, give them access to more data. Alibaba also captured the electronic payment market with the creation of Alipay and, through numerous acquisitions, entered the entertainment and social media markets as well.[108] Retailers and consumers became increasingly dependent on these infrastructures, thereby ensuring the continued supply of data about sales, payments, and user activities, among others. As Alibaba's executives have declared since 2016, Alibaba is not a retail company; it is a data company.[109] Similarly, one of Amazon's former executives, James Thomson, stated that, at Amazon, "[they] happen to sell products, but they are a data company. Each opportunity to interact with a customer is another opportunity to collect data."[110]

Both Amazon and Alibaba have developed cloud services infrastructures to enable their enterprises. Cloud computing allows platforms (of all kinds) to centralize control over data by establishing parameters for how data is produced, stored, and shared. Cloud providers (offering software- and platform-as-a-service) and data analytics firms provide specialized services on the basis of collected data, often using data extracted from the very constituencies that

(2018). Amazon's data-gathering practices have been subject to scrutiny by U.S. antitrust investigations, which revealed how Amazon uses data about retailers' value chains to expand into other sectors. STAFF OF H. SUBCOMM. ON ANTITRUST, COM. AND ADMIN. L. OF THE COMM. ON THE JUDICIARY, 116TH CONG., MAJORITY STAFF REPORT AND RECOMMENDATIONS: INVESTIGATION OF COMPETITION IN DIGITAL MARKETS 261–67 (2020) [hereinafter MAJORITY STAFF REPORT]. To ensure continuous access to consumers' behavioral data (as well as retaining and building a customer base for its own products), Amazon reportedly monitors communications between third-party marketplace merchants and consumers and penalizes those merchants who direct consumers to their own sites or other sales channels. *Id.* at 258. Amazon also convinced smaller third-party retailers to sell items via its Marketplace, offering to share with them customer analytics while retaining complete access to and control over that data, effectively "renting the Amazon customer" to third-party sellers. *Id.* at 267–73. Amazon has captured a significant (and leading) market share in the United States and is consistently increasing its market share globally. Enjoying market dominance has allowed Amazon to exercise "gatekeeping power" to cement its hold on data against competitors. *See* discussion *infra* Section II.D.

107. Khan, *supra* note 106, at 328.

108. *See* HONG SHEN, ALIBABA: INFRASTRUCTURING GLOBAL CHINA 29 (2021).

109. Alizia Staff, *Five Reasons Why Alibaba Is a Data (Not E-Commerce) Company*, ALIZILA (Oct. 17, 2016), https://www.alizila.com/five-reasons-why-alibaba-is-a-data-company/ [https://perma.cc/B5TJ-3ZZJ].

110. Kelion, *supra* note 106.

subsequently consume the companies' services.[111] Access to additional data generated through cloud computing services can also be used by the providers to train machine learning algorithms, which are subsequently offered as a service as well.[112] Although there are smaller cloud providers and a spectrum of data analytics companies, some of which focus on specific domains such as health or agricultural data, many of the big tech companies—including Alphabet, Alibaba, Baidu and Tencent—perform multiple functions, operating as platforms, cloud service providers, and data analytics companies.[113] Such integration results in concentrated control over data infrastructures and creates path dependencies for other businesses, consumers, and the public, thereby commanding "loyalty" when "exit" and "voice" are not viable.[114]

Social media giants, such as Facebook, YouTube, Twitter, and Tencent, have similarly been able to amass large stores of behavioral data, albeit through different business models. Advertising revenue dependent business models rely on platforms' abilities to create and manipulate social practices that constitute data. Platforms leverage behavioral data of users by subjecting them to granular analysis, "matching" users' desires and interests with products of companies that place targeted ads. Any subsequent interaction with that product in turn produces additional data. The ability to create, surveil, and affect "data doubles"[115] (i.e., digital representations of individuals, communities, and environments) across networked spaces and territorial borders is central to the business models for many companies, leading them to pursue corporate acquisitions of products and other data-generating platforms that could be added to expand their existing infrastructures.[116] These corporate acquisitions also

111. For examples of types of data collected by cloud service providers, see *Privacy Notice*, AMAZON WEB SERVS., https://aws.amazon.com/privacy [https://perma.cc/BK9P-Z7DY].

112. We thank Niels ten Oever for highlighting this point to us.

113. *See, e.g.*, U.S. SEC. & EXCH. COMM'N, 001-37580, FORM 10-K: ALPHABET, INC., https://www.sec.gov/Archives/edgar/data/1652044/000165204417000008/goog10-kq42016.htm [https://perma.cc/BX5U-U39D] ("Alphabet is a collection of businesses – the largest of which, of course, is Google.").

114. *See generally* ALBERT O. HIRSCHMAN, EXIT, VOICE, AND LOYALTY (1970).

115. Kevin D. Haggerty & Richard V. Ericson, *The Surveillant Assemblage*, 51 BRIT. J. SOCIO. 605, 613–14 (2000); Taylor & Broeders, *supra* note 12, at 231–32.

116. Examples include Facebook's acquisition of social media data-gathering platforms Instagram and WhatsApp, Sam Shead, *Facebook Owns the Four Most Downloaded Apps of the Decade*, BBC NEWS (Dec. 18, 2019), https://www.bbc.com/news/technology-50838013 [https://perma.cc/D5VE-VTSF], Google's acquisition of health data-gathering device FitBit,

consolidate and entrench control over data through the integration of products and services that strengthen and expand data infrastructures.

Infrastructural control over data allows companies both to extract their users' data and to *shape* what data they gather from user behavior. For example, despite content being supposedly user-generated, the extent to which content is mediated by social practices (e.g., by content moderators), manipulated through technical means (e.g., by virtue of algorithmic targeting), and removed through a combination of technical, social, and organizational means (e.g., via blocks, suspensions, and bans) places significant constraints on individuals' agency. The entire engagement can be seen as "manufactured" because users are being nudged to engage (or not) with the advertised material or product, and each iterative (non)engagement is registered as new data inputs. Such behavior-shaping power extends beyond selling products. Users are being targeted with particularized information (e.g., "suggested posts" on Instagram) deemed of potential interest on the basis of users' demographic or behavioral data. This is not just a binary matter of showing or not showing content, but also encompasses subtler forms of manipulation—for example, by shifting the order of search results and news feeds.[117]

Technical protocols and centralized control also define and structure spaces within which users can conduct their array of activities.[118] This reality is poignantly illustrated by Catherine D'Ignazio and Lauren Klein in their account of Facebook's override of user choices. In 2014, Facebook expanded the gender categories available to registered users from the conventional binary (male/female) to over fifty choices, ranging from 'Genderqueer' to 'Neither.' A year later, Facebook replaced the select-from-options model with a blank text field, a decision that was touted as being very progressive. D'Ignazio and Klein note, however, that:

Jessica Bursztynsky, *Google Closes Its Fitbit Acquisition*, CNBC (Jan. 14, 2021, 11:25 AM), https://www.cnbc.com/2021/01/14/google-closes-its-fitbit-acquisition.html [https://perma.cc/D752-DBSA], and Alibaba's acquisition of UC Browser developer UCWeb, Catherine Shu, *Alibaba Acquires UCWeb, Maker Of China's Most Popular Mobile Browser*, TECHCRUNCH (June 11, 2014, 1:22 AM), https://techcrunch.com/2014/06/10/alibaba-acquires-ucweb-maker-of-chinas-most-popular-mobile-browser [https://perma.cc/MK9T-E38E].

117. Robert Epstein & Ronald E. Robertson, *The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcome of Elections*, 112 PNAS E4512, E4512 (2015). A platform's penchant for user manipulation can have stark effects on democratic decision-making, as Jonathan Zittrain observed in *Engineering an Election*, 127 HARV. L. REV. F. 335 (2014). *See also* Evelyn Douek, *Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability*, 121 COLUM. L. REV. 759 (2021).

118. COHEN, *supra* note 14, at 42.

> [B]elow the surface, Facebook continues to resolve users' genders into a binary: either "male" or "female.". Evidently, this decision was made so that Facebook could allow its primary clients——advertisers——to more easily market to one gender or the other. Put another way, even if you can choose the gender that you show to your Facebook friends, you can't change the gender that Facebook provides to its paying customers . . . . [C]orporations like Facebook, and not individuals . . . have the power to control the terms of data collection.[119]

Control over data infrastructures is a powerful form of control over social, political, and economic organizations of human life.[120] Contesting that power is challenging. Data infrastructures exhibit a high degree of opaqueness, if not invisibility.[121] How Facebook builds user profiles or what algorithm Amazon uses to determine customers' purchasing power is not known. The datasets on which algorithms are trained and the code of internal data management software are rarely revealed. Although companies are increasingly deploying open-source software and open standards, for the most part how control over data is exercised (and by whom) in different contexts is neither apparent nor easily ascertained. Similarly, how much and what kind of data companies control might not even be known in a holistic fashion internally and is entirely inscrutable from the outside.[122] This opacity is partly purposeful (e.g., where the technology is proprietary or when concealing infrastructure provides a market advantage), partly a function of expertise and special knowledge required to understand certain components and layers of data infrastructures, and partly a function of the sheer megalomaniac size of data infrastructures that escape internal and external scrutinization due to their complexity.

Compounding the problem of contestation is that individuals' awareness of "behind the scene" control recedes due to the increasing ubiquity (and in some case addictiveness) of data-generating platforms and devices and the heightened degree to which data infrastructures

119. D'IGNAZIO & KLEIN, *supra* note 35, at 98–100. This was discovered and discussed in detail by Rena Bivens, *The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook*, 19 NEW MEDIA & SOC'Y 880 (2017).

120. DENARDIS, *supra* note 48, at 17–21.

121. FRANK PASQUALE, THE BLACK BOX SOCIETY 193 (2015).

122. See discussion *infra* Section II.C on why data protection law does not generate adequate transparency. See also discussion *infra* Section III.C on how more transparency over data could be demanded.

are being embedded and routinized in daily lives.[123] These conditions are neither incidental nor coincidental, but rather a product of deliberate design choices that render illusory any impression of choice and agency exemplified, among other things, through consent "clicks" or service agreements.[124]

Importantly, and particularly relevant for developing countries, the unequal distributive impacts of data infrastructures are a phenomenon not limited to big tech platforms. The increasing deployment of sensors for data collection, for example in digital agriculture, illustrates similar effects. Sensors in farming equipment are often linked to specific data management platforms, like Climate Field View operated by Climate Corporation.[125] Climate Corporation and other big agriculture companies, like John Deere, have invested heavily in technologies that use detailed data (e.g., on soil, seed, and weather) to provide ostensibly useful insights to farmers (e.g., on how to increase yields).[126] In practice, these investments have meant that farmers who want to benefit from data-driven farming supply data to data management platforms that are not only able to dictate the terms of service (including compensation) due to power imbalances, but also to entrench control.[127] Such control is ensconced by creating farmers' dependencies on particular data management platforms, including through proprietary software, and by limiting the interoperability of

123. *See, e.g.*, Shoshana Zuboff, *"We Make Them Dance": Surveillance Capitalism, the Rise of Instrumentarian Power, and the Threat to Human Rights*, *in* HUMAN RIGHTS IN THE AGE OF PLATFORMS 3, 17–18 (Rikke Frank Jørgensen & David Kaye eds., 2019 More generally on the power of platforms to construct realities and shape user behavior, see Mikkel Flyverbom & Glen Whelan, *Digital Transformations, Informed Realities, and Human Conduct*, *in* HUMAN RIGHTS IN THE AGE OF PLATFORMS, *supra* at 53, 53–58.

124. Critical privacy scholarship has examined the limitations of consent at length. *See generally* WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018) (arguing that technologies are designed to undermine privacy); CHRISTINE UTZ ET AL., (UN)INFORMED CONSENT: STUDYING GDPR CONSENT NOTICES IN THE FIELD 973 (2019), https://arxiv.org/pdf/1909.02638.pdf [https://perma.cc/9NPB-BZWX] (presenting empirical evidence about how users are being nudged towards giving "consent").

125. Monsanto acquired Climate Corporation for nearly $1 billion USD in 2013. Michael Specter, *Why the Climate Corporation Sold Itself to Monsanto*, NEW YORKER (Nov. 4, 2013), https://www.newyorker.com/tech/annals-of-technology/why-the-climate-corporation-sold-itself-to-monsanto [https://perma.cc/R3BZ-RCAF].

126. Sjaak Wolfert et al., *Big Data in Smart Farming – A Review*, 153 AGRIC. SYS. 69, 75 (2017).

127. Sarah Rotz et al., *Automated Pastures and the Digital Divide: How Agricultural Technologies are Shaping Labour and Rural Communities*, 68 J. RURAL STUDS. 112, 117 (2019)

sensors embedded in farming equipment with their platforms.[128] As Sarah Rotz and her colleagues have aptly summarized:

> [F]armers and farm workers continue to carry the material risks and bear the livelihood impacts of agriculture while the capital gains of digitalization are, largely, extracted by data management companies. Indeed, agricultural data have significant use value because they are an essential tool for these companies' platform and predictive algorithm development. As with capitalist modes of banking, farmers deposit their data (money) into the system. These data are then used (reinvested) by the companies to make a profit. In effect, some farmers are becoming 'digital labourers', while data management companies accumulate the economic benefits via the expansion of their knowledge systems—the new digital commodity. This is similar to the capital accumulation models of social media platforms such as Facebook and Google.[129]

As data-collecting technologies become more diffused and embedded in material systems, as their architecture becomes more complex and distributed, as path dependencies become cemented,[130] and as entities controlling them become larger and more consolidated, opportunities for the contestation of oversized control over data infrastructures, and thus of remedying of data inequalities, diminish or even vanish. In Part II of this Article, we ask about the role of law in sustaining data inequalities. Part III will thereafter inquire into potential regulatory or policy shifts that might counteract data inequality.

---

128. *Id.*

129. *Id*.

130. Bowker et al., *supra* note 5, at 97. Jean-François Blanchette notes:

> [Computing infrastructure] developed incrementally, from the progressive laying down of its infrastructural components, including data centers, fiber cables, economic models, and regulatory frameworks. Such incremental development means that early-stage design choices *persist*, often with unforeseen consequences and become increasingly difficult to correct as the infrastructure becomes ubiquitous, its functionality expands, and the nature of the traffic it serves evolves . . . .

Jean-François Blanchette, *Introduction: Computing's Infrastructural Moment*, *in* REGULATING THE CLOUD: POLICY FOR COMPUTING INFRASTRUCTURE 1, 3 (Christopher S. Yoo & Jean-François Blanchette eds., 2015).

II. LEGAL DIMENSIONS OF DATA INEQUALITY

Extant law does not effectively address data inequality enabled by concentrated control over the data infrastructures identified in Part I. At best, law tends to ignore the infrastructural root causes of data inequality. At worst, it may contribute to or even entrench data inequality. In this Part, we explore the legal dimensions of data inequality to substantiate these claims. Our analysis straddles several established domains of law that pertain to relevant components of data infrastructures, with a focus on: (1) data protection and privacy law; (2) intellectual property law; and (3) antitrust and competition law. We cannot, however, offer comprehensive legal analysis in this Article. For example, we bracket the important and intricate tax law questions raised by globally operating corporations' tax avoidance strategies,[131] even though some claim that they are particularly pronounced in the digital economy.[132] We also do not explore corporate law and corporate governance interventions, despite their critical importance to the ways in which multinational corporations can exercise control over complex and transnationally diffused data infrastructures.[133]

---

131. On the evidence about tax avoidance, see generally Nadine Riedel, *Quantifying International Tax Avoidance: A Review of the Academic Literature*, 69 REV. OF ECON. 169 (2018); Robert Bird & Karie Davis-Nozemack, *Tax Avoidance as a Sustainability Problem*, 151 J. BUS. ETHICS 1009 (2018). For Latin American and Caribbean regional perspectives on taxation and the digital economy, see the various contributions in the online symposium convened by Monica Victor at AfronomicsLaw. *Symposia: Forthcoming Feature Symposium*, AFRONOMICSLAW, https://www.afronomicslaw.org/symposia/ [https://perma.cc/7UV5-5MJ8]. On the OECD's multilateral initiatives and their impact on international tax law, see generally Ruth Mason, *The Transformation of International Tax*, 114 AM. J. INT'L L. 353 (2020).

132. *See, e.g.*, Grant Richardson & Grantley Taylor, *Income Shifting Incentives and Tax Haven Utilization: Evidence from Multinational U.S. Firms*, 50 INT'L J. ACCT. 458 (2015) (finding that multinationality, transfer pricing aggressiveness, thin capitalization, and intangible assets are positively associated with tax haven utilization). In August 2016, the European Commission took action against Ireland for enabling such tax avoidance strategies, framing its arrangements with Apple as illegal state aid amounting to EUR 13 billion in unlawful tax advantages. In July 2020, the EU's General Court annulled this decision in Joined Cases T-778/16 and T-892/16, Ireland v. Commission, ECLI:EU:T:2020:338, ¶¶ 505–07 (July 15, 2020). The European Commission has appealed the judgment to the EU's Court of Justice.

133. *See* discussion *supra* Section I.C; *see also* John Ruggie, *Multinationals as Global Institution: Power, Authority and Relative Autonomy,* 12 REGUL. & GOVERNANCE 317, 317 (2018) (contrasting corporate social responsibility with the imposition of binding legal obligations on multinational enterprises).

Our analysis integrates domestic and international law because separation between these two "levels" is ultimately artificial, especially in the digital domain.[134] Throughout, we highlight the growing importance of international economic law, which has acquired characteristics of "megaregulation" in recent "comprehensive" trade and investment agreements.[135] These agreements increasingly create important secondary rules that shape and often constrain states' abilities to regulate data flows, ownership, protection, and competition.[136]

Our aim is to illustrate how various fields of private and public law affect uneven control over data and data infrastructures. In doing so, we are inspired by the analysis of the legal coding of capital that Katharina Pistor has pioneered and extended to the emerging legal coding of data.[137] We also build upon the critical accounts by Julie Cohen and Amy Kapczynski about the ways in which platform companies have used legal technologies to advance their interests in an information-capitalist economy.[138] Our goal is not to blame "the law" for data inequality, but rather to render visible how different domains of law are entangled with data infrastructures and to demonstrate how lawyers can use different "legal technologies" to facilitate corporate control over data.[139] We certainly do not think that the many scholars, practitioners, and activists who are engaged in the legal domains of intellectual property, data protection and privacy, or

134. On the 'marbled' structure of global governance, see Sabino Cassese, *Governing the World*, *in* RESEARCH HANDBOOK ON GLOBAL ADMINISTRATIVE LAW 502, 506–10 (Sabino Cassese ed., 2016).

135. *See* Benedict Kingsbury, Paul Mertenskötter, Richard B. Stewart & Thomas Streinz, *TPP as Megaregulation*, *in* MEGAREGULATION CONTESTED: GLOBAL ECONOMIC ORDERING AFTER TPP ch. 2 (Benedict Kingsbury et al eds., 2019). *But see* HANDBOOK OF DEEP TRADE AGREEMENTS (Aaditya Mattoo, Madia Rocha & Michele Ruta eds., 2020) (conceptualizing "deep trade agreements" as trade agreements that branch out into additional policy areas in pursuit of "deep integration").

136. *See* Thomas Streinz, *Digital Megaregulation Uncontested? TPP's Model for the Global Digital Economy,* MEGAREGULATION CONTESTED: GLOBAL ECONOMIC ORDERING AFTER TPP, *supra* note 135, at 312. *See generally* Svetlana Yakovleva & Kristina Irion, *The Best of Both Worlds? Free Trade in Services, and EU Law on Privacy and Data Protection*, 2 EUR. DATA PROT. L. REV. 191 (2016).

137. *See generally* PISTOR, *supra* note 10; Katharina Pistor, *Rule by Data: The End of Markets?*, 83 L. & CONTEMP. PROBS. 101 (2020).

138. *See generally* COHEN, *supra* note 14; Kapczynski, *supra* note 14.

139. We adopt the "legal technologies" term from Davis, *supra* note 61, at 83, who defines technology as useful knowledge about how to produce things at low cost (studying contract law innovations). We depart from the focus on innovations and instead highlight how lawyers' knowledge and skill can be used in data contracting, licensing, and property claims.

competition law are not addressing important issues. Our argument is much narrower: These domains of law are often not attuned to infrastructural control over data and therefore do not effectively regulate data inequality (at best) or might entrench it (at worst). This is also not to say that all these domains of law need to change to confront data inequality. As we explore in Part III below, remedying data inequality requires carefully-calibrated interventions within the law and beyond.

In the following Sections, we first highlight how the legal system has facilitated the global "free flow" of data through the internet, which crucially relies on physical infrastructures and interoperability standards, without much concern as to between whom data flows and where it accumulates.[140] We then turn to the much-discussed question of legal data ownership.[141] We find that even where the law recognizes no property rights in data, control over data is achieved through control of the relevant data infrastructures encoded in contractual terms, irrespective of formal property rights. However, when this status quo is challenged by demands for transparency or data sharing, lawyers are likely to invoke property or property-like rights in data on behalf of data holders who will lobby for the recognition of such rights. Next, we address the dominant approach to contemporary data regulation: rights-based data protection and privacy law.[142] While we recognize the importance of this domain and the global diffusion of data protection and privacy rights, we ultimately conclude that this rights-based approach does not effectively confront data inequality. At best, it raises the costs of data accumulation, but it has not, at least thus far, effectively curtailed data hoarding. At worst, this approach privileges those with the means to shoulder increased compliance costs, thereby inadvertently exacerbating data concentration. Finally, we discuss the law applicable to platform companies.[143] We recognize the evolving debate about whether and how antitrust and competition law should confront platform power, and we concede that this approach is more attuned to dynamics of infrastructural control than the other areas of law that we explore. But we also caution that antitrust and competition law come with certain assumptions about market efficiencies and consumer welfare that can make them blind to broader concerns around data inequality, which ought to be addressed through other means.

140. *See* discussion *infra* Section II.A.

141. *See* discussion *infra* Section II.B.

142. *See* discussion *infra* Section II.C.

143. *See* discussion *infra* Section II.D.

A. "Free Flow" of Data

The global expansion of the internet since the early 1990s has enabled an unprecedented degree of inter-connectedness of communication networks and devices.[144] This development has often been hailed as promoting individual and collective freedom globally,[145] despite persistent "digital divides" between and within countries.[146] While first commonly perceived as a communication infrastructure, the internet has since evolved into an indispensable infrastructure for the generation, processing, and transfer of data more generally.[147] Internet governance has been subject to increasing contestation for a variety of reasons. The traditional preponderance of U.S. stakeholders in internet governance institution has been increasingly challenged.[148] Tensions among multistakeholderism, intergovernmentalism, and nation states' jurisdiction remain unresolved.[149] The increasing pushback against the global dominance

144. *See, e.g.*, RICHARD BALDWIN, THE GREAT CONVERGENCE: INFORMATION TECHNOLOGY AND THE NEW GLOBALIZATION 81–84 (2016) (discussing the ICT revolution).

145. *See* Ira C. Magaziner, *Creating a Framework for Global Electronic Commerce, Future Insight*, Release 6.1 PROGRESS & FREEDOM FOUNDATION (July 1999), http://www.pff.org/issues-pubs/futureinsights/fi6.1globaleconomiccommerce.html [https://perma.cc/7CV9-JXJB] (claiming that the internet was promoting "individual freedom and individual empowerment" and that it would "bring all the peoples of the world closer together"). *But see* David Pozen, *The De-Americanization of Internet Freedom*, LAWFARE (June 13, 2018, 2:19 PM), https://www.lawfareblog.com/de-americanization-internet-freedom [https://perma.cc/MR7D-X2NK]; Jack Goldsmith, *The Failure of Internet Freedom*, KNIGHT FIRST AMEND. INST. AT COLUM. UNIV. (June 13, 2018), https://knightcolumbia.org/content/failure-internet-freedom [https://perma.cc/4TVG-DZ3L].

146. *See generally* JAN VAN DIJK, THE DIGITAL DIVIDE (2020). *See* discussion *supra* Section I.B on different forms of data inequality.

147. *See* discussion *supra* Section I.C. *See generally* DENARDIS, *supra* note 48 (describing the internet's transformation from a communication into a control network).

148. *See generally* ANDREW L. RUSSELL, OPEN STANDARDS AND THE DIGITAL AGE: HISTORY, IDEOLOGY, AND NETWORKS (2014); MILTON L. MUELLER, NETWORKS AND STATES: THE GLOBAL POLITICS OF INTERNET GOVERNANCE (2010); LAURA DENARDIS, THE GLOBAL WAR FOR INTERNET GOVERNANCE (2014).

149. *See generally* Mark Raymond & Laura DeNardis, *Multistakeholderism: Anatomy of an Inchoate Global Institution*, 7 INTERNATIONAL THEORY 572 (2015) (contrasting multistakeholderism with multilateralism); Niels ten Oever, *The Metagovernance of Internet Governance, in* POWER AND AUTHORITY IN INTERNET GOVERNANCE: RETURN OF THE STATE? 56 (B. Haggart et al, eds., 2021) (contrasting the private and multistakeholder internet governance regime and the multilateral internet governance regime); Thomas Streinz, *Global Hybrid Internet Governance: The Internet Corporation for Assigned Names and Numbers (ICANN)*, *in* Global Hybrid and Private Governance: Standard-Setting, Market Regulation, and Institutional Design (manuscript at 91–97) (Benedict Kingsbury & Richard B. Stewart

of U.S.- and China-based corporations and the varieties of "surveillance capitalism" that they orchestrate is spilling over into internet governance.[150] At the same time, rising geopolitical tensions between the United States and China complicate engagement in multilateral and multistakeholder settings and call into question whether "free flow" of data will remain a viable default position.

The U.S. government fostered a shift from public funding and public management towards commercialization and privatization when it decided to facilitate the economic exploitation of the internet during the early 1990s.[151] Ever since, private sector leadership of the internet has been a staple and mantra of internet governance.[152]

The internet has been remarkably successful in facilitating data transfers across borders. The relevance of data mobility for the unbundling of economic production—creating transnational networks in which goods and services are being exchanged and recombined, often within firms, seems beyond doubt, yet remains difficult to quantify.[153] The internet's ability to route information through inter-connected networks depends on physical connectivity (either through a cable or via the electromagnetic spectrum), interoperable protocols that govern the exchange of data between different networks (the proverbial inter-networking), and the absence of legal limitations

---

eds., 2022) (analyzing the role of states within internet governance institutions and the continued relevance of national jurisdictions).

150. *See* ZUBOFF, *supra* note 33 (arguing that Google and Facebook engineered "surveillance capitalism" with detrimental societal effects in the U.S.); Brett Aho & Roberta Duffield, *Beyond Surveillance Capitalism: Privacy, Regulation and Big Data in Europe and China*, 49 ECON. & SOC'Y 187, 188–89 (2020) (arguing that the EU and China are trying to move beyond surveillance capitalism, albeit in different ways). On the geoeconomics and geopolitical interplay between corporations and governments engaged in surveillance activities, see Madison Cartwright, *Internationalising State Power Through the Internet: Google, Huawei and Geopolitical struggle*, 9 INTERNET POL'Y REV. 1, 8–12 (2020).

151. *See* Brett Frischmann, *Privatization and Commercialization of the Internet Infrastructure: Rethinking Market Intervention into Government and Government Intervention into the Market*, 2 COLUM. SCI. & TECH. L. REV. 1, 7 (2001).

152. *See* Madeline Carr, *Power Plays in Global Internet Governance*, 43(2) MILLENNIUM 640, 645–47 (2015). *See generally* Kal Raustiala, Editorial Comment, *Governing the Internet*, 110 AM. J. INT'L L. 491 (discussing the transition from U.S. governmental oversight). Not creating a government-led or intergovernmental governance structure in its place was one of the conditions the U.S. National Telecommunications and Information Administration (NTIA) postulated before giving up oversight over key administrative functions at the heart of the Internet's domain name system. *Id.* at 499.

153. *See generally* BALDWIN, *supra* note 144; Milton Mueller & Karl Grindal, *Data Flows and the Digital Economy: Information as a Mobile Factor of Production*, 21 DIGIT. POL'Y, REGUL. & GOVERNANCE 71 (2019).

imposed by governments or private actors with an interest in blocking, monitoring, or at least limiting cross-border data movements.[154]

Below, we illustrate how law has interacted with different components of internet infrastructure to enable global data flows, while also creating, sustaining, or at least ignoring inequalities in terms of data access, control, and governance.

First, we highlight how the international law of the sea has accommodated the colonial path-dependencies of submarine cables. Today, these cables are often owned and operated by the same actors that enjoy concentrated control over complex data infrastructures.[155] We then illustrate how protocols enable interoperability between networks, while social norms within global internet standard-setting bodies limit private sector ability to avail themselves of intellectual property rights, thereby preserving inter-connectivity without allowing much space for countervailing interests.[156] We next expose how emerging rules of international economic law protect transnational data mobility against governmental control in the form of cross-border data transfer limitations and territorial data localization requirements, thereby limiting states' abilities to counteract data inequality in this way.[157] Lastly, we conclude with examples that illustrate evolving internet-networking dynamics and power struggles on the world wide web ("www"), shining a light on the power differentials between different networks on the internet, as mediated through contracting or private norms. The emergence of content delivery networks ("CDNs"), in particular, affects data flows in ways that seem in tension with established net neutrality principles without violating net neutrality laws.[158]

The internet can be modeled as consisting of several layers together forming a "stack," with each layer facilitating a discrete function, on which layers above rely.[159] At the bottom, and hence most fundamental, is the need to establish a physical network connection. The internet's backbone for cross-border data flows consists of a

154. *See* discussion *infra* II.A.

155. *See* discussion *infra* notes 160–167 and accompanying text.

156. *See* discussion *infra* notes 169–180 and accompanying text.

157. *See* discussion *infra* notes 181–199 and accompanying text.

158. *See* discussion *infra* notes 200–212 and accompanying text.

159. There are different models of the internet stack; for a discussion of the historical debate, see generally Russell, *supra* note 148 (contrasting the complex Open Systems Interconnection model (OSI model) with the ultimately successful Internet Protocol Suite). *See also* YOCHAI BENKLER, THE WEALTH OF NETWORKS ch. 11 (2006) (distinguishing between physical, logical, and content layers to identify mechanisms of openness and enclosure).

global network of submarine fiber-optic cables, laid on the seabed.[160] Imperialism, free trade policies, and state subsidies shaped the locations and ownership of early submarine cables (used for telegraph services), connecting British, European, Japanese, and American empires to their colonies and dominions overseas.[161] These routes, laid down in the nineteenth century, remain the most important corridors for modern fiber-optic cables.[162] The extent to which states are able to regulate the surveillance, construction, maintenance, and use of modern fiber-optic cables is guided to some extent by the international law of the sea.[163]

Uneven access to transnational connectivity and disparate control over submarine cables are largely functions of historical legacies and contemporary market dynamics. The centrality of national sovereignty and territorial jurisdiction in the regulatory regime for submarine cables privileges coastal (and archipelagic) states and states that host land cable routes. States dependent on others for internet connectivity are at a natural disadvantage. The relative power balance between states and cable operators mediates their respective legal rights and de facto control over physical internet infrastructures.[164] Although government ownership and financing of

160. *See generally* NICOLE STAROSIELSKI, THE UNDERSEA NETWORK (2015); SUBMARINE CABLES: THE HANDBOOK OF LAW AND POLICY 2 (Douglas R. Burnett, Robert C. Beckman & Tara Davenport eds., 2014). The global cable network is estimated to comprise more than 400 separate submarine cable systems, stretching over 1.2 million kilometres and carrying the bulk of global internet traffic (estimated at 97% in 2014). *See* TeleGeography, SUBMARINE CABLE MAP, https://www.submarinecablemap.com [https://perma.cc/4QWW-JFFG].

161. *See generally* Roxana Vatanparast, *The Infrastructures of the Global Data Economy: Undersea Cables and International Law*, 61 HARV. INT'L L.J. FRONTIERS 1 (2020).

162. Dwayne Winseck, *The Geopolitical Economy of the Global Internet Infrastructure*, 7 J. INFO. POL'Y 228, 237 (2017) ("Communication paths . . . link many of the same 'world cities' now as they did then and some of the same old ornate cable telegraph buildings of the nineteenth century in London and New York have even been retrofitted for fiber optic cables today.").

163. The U.N. Convention on the Law of the Sea (UNCLOS) recognizes sovereign control of a coastal state over a twelve-nautical-mile belt of sea (i.e., its territorial sea), which includes the airspace above and the seabed and subsoil below. U.N. Convention on the Law of the Sea arts. 2, 3, Dec. 10, 1982, 1833 U.N.T.S. 397. An archipelagic state has sovereignty over the waters enclosed by its archipelagic baselines (archipelagic waters) with the express obligation to respect existing submarine cables laid by other states and to permit maintenance and replacement of such cables on receiving due notice. *Id.* art. 51(2).

164. *See* Winseck, *supra* note 162, at 236 (2017):

> The monopoly landing rights that [states] typically gave in the early years of development varied considerably, as did the terms of service they demanded with respect to privileges to be provided to local state officials and interconnection with local telegraphs, as well as their need to monitor (surveillance) and block

fiber-optic cables are both on the rise, they often use public-private partnerships, with investments made by telecommunications operators, non-telecommunication companies with high capacity demands for their private networks, and investment banks.[165] Internet platform companies have joined traditional telecommunications companies to become cable co-owners in consortia, reflecting their growing power in internet governance as they increase their control over the scarce resource of transnational bandwidth.[166] Despite these trends, relative ownership and control over core elements of the global internet infrastructure is shifting away from U.S. firms towards new internet infrastructure providers located in emerging markets, especially in Asia.[167]

Building on the physical link layer, in the next layer names and numbers are assigned to the nodes of the network to make them uniquely identifiable, and protocols coordinate the transport of data between them. Inter-networking requires interoperable standards that are being developed by various standard-setting organizations, such as the Internet Engineering Task Force (IETF), the Internet Corporation

---

(censorship) messages perceived as threats to public morality or national security. These landing licenses typically reflected the strength of the state that negotiated them. The stronger the state, the less likely it was to grant monopoly rights, as was the case in Britain and the United States, whereas the weaker the state, the longer the right to a monopoly, the more restrictive the terms of service obligations, and the less likely companies were to cooperate in ways other than those that advanced their business interests.

165. *See* World Bank, 2017 ICT BACKBONE SECTOR: PRIVATE PARTICIPATION IN INFRASTRUCTURE (PPI), https://ppi.worldbank.org/en/ppi#2 [https://perma.cc/S2CM-MLHD] (surveying ICT-backbone infrastructure projects, including land-based and submarine cables, with an active government component).

166. *See* Mozilla, *The New Investors in Underwater Sea Cables*, INTERNET HEALTH REP. 2019 (Apr. 2019), https://internethealthreport.org/2019/the-new-investors-in-underwater-sea-cables [https://perma.cc/5VZP-8QBR] (finding that Google, Facebook, Amazon, and Microsoft owned or leased more than half of the undersea bandwidth in 2018); *see also* Bikash Koley, *Hola, South America! Announcing the Firmina Subsea Cable*, GOOGLE CLOUD BLOG (June 9, 2021), https://cloud.google.com/blog/products/infrastructure/announcing-the-firmina-subsea-cable [https://perma.cc/NJ2Y-DLBZ] (touting a new subsea cable project between the East Coast of the United States and Argentina with additional landings in Brazil and Uruguay); Tage Kene-Okafor, *Facebook-Backed 2Africa Set to Be the Longest Subsea Cable Upon Completion*, TECHCRUNCH (Sept. 29, 2021, 8:27 PM), https://techcrunch.com/2021/09/29/facebook-backed-2africa-set-to-be-the-longest-subsea-cable-upon-completion [https://perma.cc/JW92-ARQG]. On submarine cable projects orchestrated by Chinese infrastructure companies, see Matthew S. Erie & Thomas Streinz, *The Beijing Effect: China's Digital Silk Road as Transnational Data Governance*, 54 N.Y.U. J. INT'L L. & POL. 1, 50–52 (2022).

167. Dwayne Winseck, *Internet Infrastructure and the Persistent Myth of U.S. Hegemony*, *in* INFORMATION, TECHNOLOGY AND CONTROL IN A CHANGING WORLD 93, 101–15 (B. Haggart et al. eds., 2019).

for Assigned Names and Numbers (ICANN), or the World Wide Web Consortium (W3C).[168] Illustrations of the internet stack—such as the following—often list key standards and relevant standard-setting organizations, but fail to account for the data infrastructures built on top of the internet's application layer. They also tend to conceal the identities of dominant stakeholders within standard-setting organizations.[169]

Table 1[170]

Layer	**Function**	**Standards**	**Organizations**
Application	Email, WWW	e.g., IMAP, HTTP/HTML	W3C, IETF
Transport	Transfer	e.g., TCP, UDP	IETF
Internet	Addressing and Routing	e.g., IP, BGP	ICANN, IETF
Network	Physical Connectivity	e.g., Ethernet, Wi-Fi, 4G/LTE	IEEE

The internet's open standards enable everyone to connect as long as the participants' machines comply with the protocols.[171] In other words, standards "regulate" how data "flows" on the internet.[172] This reality turns ostensibly-technical standard-setting organizations, such as the IETF, into global regulators of data flows, raising

168. *See* Jorge L. Contreras, *Patents and Internet Standards*, GCIG Paper No. 29, GLOBAL COMMISSION ON INTERNET GOVERNANCE PAPER SERIES 2 (Apr. 15, 2016), https://www.cigionline.org/static/documents/gcig_no29_web.pdf [https://perma.cc/6PVE-BBYV].

169. *See, e.g.*, *id.* at 2 tbl.1.

170. *Id.*

171. *See generally* RUSSELL, *supra* note 148 (contrasting standards for previous telecommunication networks with the standards that enabled the internet).

172. The scare quotes around "regulate" are meant to indicate that standards are formally voluntary but often de facto unavoidable, whereas governmental regulation is formally binding but not always complied with. On the complex relationship between formal law and technical standards, see Benedict Kingsbury, *Preface*, *in* CAMBRIDGE HANDBOOK OF TECHNICAL STANDARDIZATION LAW (VOL. 2): FURTHER INTERSECTIONS OF PUBLIC AND PRIVATE LAW at xv (Jorge L. Contreras ed., 2019). The scare quotes around "flows" echo our discussion of data metaphors above, see discussion *supra* Section I.A, which emphasized that data does not move without agency but is being sent and received.

inevitable questions about interests and politics.[173] These conflicts have materialized in attempts to integrate international human rights law as a substantive standard, or at least a discursive toolkit, to guide their decision-making.[174] As it turns out, internet standard-setting organizations are so committed to global connectivity that they perceive any interests that are in tension with this goal as irritants.[175]

Their commitment to connectivity explains why the internet's core standard-setting organizations have been remarkably successful in preventing encroachment by intellectual property law. Technologies manufactured in accordance with the protocols and parameters specified by standards can, in principle, enjoy patent protection.[176] Complex products may implement dozens or even hundreds of standards, each of which may in turn be covered by numerous standards-essential patents ("SEPs").[177] Most patents are typically owned by firms themselves engaged in the standards-development process, thus making governance structures of the standards-setting organizations and the opportunity to participate in them an important site of infrastructural control.[178] In the context of internet-related standards, however, the dominant standard-setting organizations have developed policies and norms requiring the

173. *See generally, e.g.*, Corinne Cath & Luciano Floridi, *The Design of the Internet's Architecture by the Internet Engineering Task Force (IETF) and Human Rights*, 23 SCI. & ENG'G ETHICS 449 (2017).

174. See the work by the IETF's Human Rights Protocol Consideration Research Group, which is tasked with researching whether internet standards and protocols can enable, strengthen or threaten human rights generally, albeit with a focus on traditional civil and political rights. *Human Rights Protocol Considerations*, DATATRACKER, https://datatracker.ietf.org/rg/hrpc/about/ [https://perma.cc/B7E2-RYHL]. *See generally also* Monika Zalnierute & Stefania Milan, *Internet Architecture and Human Rights: Beyond the Human Rights Gap*, 11 POL'Y & INTERNET 6 (2019); Monika Zalnieriute, *Human Rights Rhetoric in Global Internet Governance: New ICANN Bylaw on Human Rights*, 10 HARV. BUS. L. REV. 1 (2020).

175. *See generally* NIELS TEN OEVER, WIRED NORMS: INSCRIPTION, RESISTANCE, AND SUBVERSION IN THE GOVERNANCE OF THE INTERNET INFRASTRUCTURE (2020).

176. Contreras, *supra* note 168, at 2–3.

177. *Id.* SEPs are patents that will always be infringed by a product conforming to a particular standard. The existence of patents covering standards has, in some cases, led to patent wars, royalty stacking (which makes it prohibitively expensive for competitors to develop standard-complying products), and patent hold-ups (instances where SEP holder demands excessive royalties after product manufacturers have made significant investments in standardized technology, thus resulting in lock-in effects). *Id.* at 2.

178. *See* Panos Delimatsis, Olia Kanevskaia Whitaker & Zuno Verghese, *Exit, Voice and Loyalty: Strategic Behavior in Standards Development Organizations*, TILEC Discussion Paper No. DP 2019-022 (Dec. 2, 2019), https://dx.doi.org/10.2139/ssrn.3487466 [https://perma.cc/E2DW-F98K].

licensing of relevant patents on a royalty-free basis, treating standards as a type of public good that should benefit everybody without restrictions, or at least at rates that are "fair, reasonable, and nondiscriminatory."[179] While participation in nominally "global" internet governance institutions remains uneven and is dominated by certain actors—mainly company representatives and academics who can afford attendance—they have succeeded in maintaining globally uniform standards that are "open" for anyone to adopt to enable transnational connectivity.[180]

At the same time, for reasons of institutional design and ideology, internet governance institutions have largely refrained from critically examining the distributive outcomes and power dynamics that the internet has enabled as it transformed into a foundational infrastructure for data creation, processing, and transfers. In an arrangement resembling the tenuous balance between transnational economic integration and domestic societal safeguards that John Ruggie has dubbed "embedded liberalism",[181] nation states remain primarily responsible for the well-being of their citizens in the internet era. At the same time, and in contrast to prior telecommunication technology (from the telegraph to the telephone), states do not enjoy a comparable level of control over the institutions that control internet infrastructure.[182] States may, however, resort to measures that limit the cross-border transfer of data, thereby challenging the internet's foundational logic and most celebrated achievement for a variety of reasons, ranging from the pursuit of societal objectives such as data

179. See, for example, the Open Stand initiative, supported by IEEE, the Internet Society (ISOC), the Internet Engineering Task Force (IETF), the Internet Architecture Board (IAB), and W3C, and other standard-setting organizations, which provides that "Affirming standards organizations have defined procedures to develop specifications that can be implemented under fair terms. Given market diversity, *fair terms may vary from royalty-free to fair, reasonable, and non-discriminatory terms (FRAND)*." *Principles*, OPEN STAND, https://open-stand.org/about-us/principles [https://perma.cc/6TYS-DEVE] (emphasis added). For an in-depth discussion of patents, internet and standards, see generally Jorge L. Contreras, *A Tale of Two Layers: Patents, Standards and the Internet*, 93 DENV. L. REV. 855 (2016).

180. *See* BENKLER, *supra* note 159, ch. 11 (discussing standard-setting as a mechanism for openness).

181. *See* Rawi Abdelal & John G. Ruggie, *The Principles of Embedded Liberalism: Social Legitimacy and Global Capitalism*, *in* NEW PERSPECTIVES ON REGULATION 151, 151–62 (David Moss & John Cisternino eds., 2009).

182. *See generally* MUELLER, *supra* note 148 (arguing against the assumption that the internet will inevitably be subordinated to nation states). *But see* Dan Drezner, *The Global Governance of the Internet: Bringing the State Back In*, 119 POL. SCI. Q. 477, 478 (2004) (arguing that powerful states remain the primary actors for handling the internet's social and economic externalities).

protection or economic welfare, to national security concerns or political self-preservation qua censorship.[183]

One prominent example of such measures is the European Union's General Data Protection Regulation (GDPR) that limits the cross-border transfer of personal data by default.[184] The most coveted way to overcome this limitation is the "adequacy assessment" under which the European Commission determines whether another jurisdiction provides for an "essentially equivalent" level of protection.[185] The European Union's Court of Justice (ECJ) later extended this requirement to other legal technologies, especially to standardized contractual clauses that are available to "export" personal data from the European Union to other jurisdictions.[186] The GDPR hence discriminates between jurisdictions with "essentially equivalent" data protection laws and those without, and between entities that are able to provide prerequisite "additional safeguards" and those who are not.[187] The backlog that the European Commission has accumulated—and discrepancies between countries that received an adequacy finding in the past and those that arguably provide a more robust level of data protection without being granted that status—

183. The extensive literature on such restrictions often worries about "internet fragmentation" or a "splinternet." *See generally, e.g.*, MILTON MUELLER, WILL THE INTERNET FRAGMENT? (2017); Mark Lemley, *The Splinternet*, 70 DUKE L.J. 1397 (2021). *See also* Daniel Lambach, *The Territorialization of Cyberspace*, 22 INT'L STUD. REV. 482 (2020) (describing how different actors—not just states—territorialize and reterritorialize "cyberspace"); Niels ten Oever, *The Metagovernance of Internet Governance*, *in* CONTESTED POWER AND AUTHORITY IN INTERNET GOVERNANCE: RETURN OF THE STATE? ch. 3 (Blayne Haggart, Natasha Tusikov & Jan Aart Scholte eds., 2021) (differentiating between a private and multistakeholder internet governance regime and a multilateral internet governance regime, the latter of which seeks to accommodate national and regional norms and values).

184. Regulation 2016/679 of Apr. 27, 2016, On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119/1) art. 44 [hereinafter GDPR]. This restriction on cross-border transfers of personal data goes back to the 1995 Data Protection Directive and even earlier national data protection laws. *See* Paul M. Schwartz, *European Data Protection Law and Restrictions on International Data Flows*, 80 IOWA L. REV. 472, 484 (1995); CHRISTOPHER KUNER, TRANSBORDER DATA FLOWS AND DATA PRIVACY LAW 40–49 (2013).

185. This standard was clarified by the European Union's Court of Justice (ECJ) in Case C-362/14, Maximillian Schrems v. Data Protection Commissioner (Schrems), ECLI:EU:C:2015:650 (Oct. 6, 2015).

186. Case C-311/18, Data Protection Commissioner v. Facebook Ireland Ltd. & Maximillian Schrems (Schrems II), ECLI:EU:C:2020:559 (July 16, 2020).

187. Some entities might accordingly choose not to transfer personal data from the European Union at all. *See* Anupam Chander, *Is Data Localization a Solution for* Schrems II*?*, 23 J. INT'L ECON. L. 771, 771 (2020).

raises questions about the European Union's compliance with its non-discrimination commitments and applicable regulatory disciplines under the law of the World Trade Organization (WTO), in particular the General Agreement on Trade in Services (GATS).[188] The European Union's trade and internet policy confronts a fundamental tension between the GDPR's restrictions on transfers of personal data and the commitment towards unimpeded "data flows" that animates many internet governance institutions and the U.S. "digital trade" agenda.[189]

India is another prominent jurisdiction that has experimented with a variety of data localization requirements and data transfer restrictions.[190] Some of these obligations are sectoral and ostensibly motivated by safety and security concerns. The Reserve Bank of India, for example, requires payment service providers to store data relating to payment systems "only in India."[191] India's protracted effort to reform its data protection law featured different forms of data localization in subsequent drafts of its data protection bills, which are applicable to different categories of personal data.[192] In February 2019, India published its draft e-commerce policy, which adopts a nationalized version of the "data as a resource" framing and views data localization requirements with regard to certain categories of data as an important instrument in retaining data-generated value within

188. Svetlana Yakovleva & Kristina Irion, *Towards Compatibility of the EU Trade Policy with the General Data Protection Regulation*, 114 AJIL UNBOUND 10, 11 (2020). On potential justifications, see generally Neha Mishra, *Privacy, Cybersecurity, and GATS Article XIV: A New Frontier for Trade and Internet Regulation?*, 19 WORLD TRADE REV. 341 (2020).

189. *See generally* Svetlana Yakovleva & Kristina Irion, *Pitching Trade Against Privacy: Reconciling EU Governance of Personal Data Flows with External Trade*, 10 INT'L DATA PRIV. L. 201 (2020).

190. *See* ARIDRAJIT BASU, ELONNAI HICKOCK & ADITYA SINGH CHAWLA, THE LOCALISATION GAMBIT: UNPACKING POLICY MEASURES FOR SOVEREIGN CONTROL OF DATA IN INDIA, CTR. FOR INTERNET & SOC'Y 9 (2010); Aridrajit Basu, *The Retreat of the Data Localization Brigade: India, Indonesia and Vietnam*, DIPLOMAT (Jan. 10, 2020), https://thediplomat.com/2020/01/the-retreat-of-the-data-localization-brigade-india-indonesia-and-vietnam [https://perma.cc/YL22-MGFK]; *see also* Lok Sabha, Draft Report of the Joint Committee on the Personal Data Protection Bill, 2019, at 9–11 (issued Dec. 16, 2021) (discussing types, objectives, and stakeholders of data localization).

191. *See* Reserve Bank of India, Directive on Storage of Payment System Data, RBI/2017-18/153 (issued Apr. 6, 2018), https://src.bna.com/D5n?_ga=2.123343947.1131408265.1575749815-1922557974.1575749815 [https://perma.cc/B85X-XAGB]; Payment and Settlement Systems Act, 2007, § 18 (India).

192. Draft Personal Data Protection Bill, 2018, § 40 (India); Draft Personal Data Protection Bill, 2019, Bill No. 373 of 2019 § 33 (India); *see also* Draft Report of the Joint Committee on the Personal Data Protection Bill, *supra* note 190.

India.[193] In this way, India embraces an openly data protectionist approach to digital development that challenges the "free flow of data" paradigm celebrated by internet governance institutions and most trade economists.[194]

To preserve "free data flows" against governmental interference, the United States developed a new model of rules for the digital economy during the negotiations for the Trans-Pacific Partnership (TPP).[195] Even though the Trump Administration ultimately withdrew the United States from the TPP, the TPP's rules on the "free flow" of data are now in effect through the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP).[196] Subsequent agreements, including the new U.S.-Mexico-Canada Agreement (USMCA), the U.S.-Japan Digital Trade Agreement (USJDTA), and the Digital Economy Partnership Agreement (DEPA) between Chile, Singapore, and New Zealand, all adopted essentially the same model.[197] These commitments to free data flows materialize not only between countries that sign on to these agreements, but also with respect to multinational corporations easily availing themselves of a corporate nationality to protect them from "unnecessary" data transfer limitations and data localization requirements. While beneficial for the global preservation of free data flows, "free data flow" obligations deprive countries of alternative digital development models that might favor their homegrown digital economy, as India tries to do. They also complicate differentiated approaches under which only data flows to certain jurisdictions are allowed, as in the case of the European Union's regime for outward transfers of personal data.[198]

---

193. Draft National e-Commerce Policy: India's Data for India's Development, https://dpiit.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf [https://perma.cc/5FPC-M4PJ].

194. *See* Neha Mishra, *Data Governance and Digital Trade in India: Losing Sight of the Forest for the Trees?*, *in* DATA SOVEREIGNTY ALONG THE DIGITAL SILK ROAD (forthcoming), https://papersssrn.com/sol3/papers.cfm?abstract_id=3835497 [https://perma.cc/3AXE-9UTD].

195. Streinz, *supra* note 136, at 314.

196. The CPTPP entered into force for Australia, Canada, Japan, Mexico, New Zealand, and Singapore in December 2018, and for Vietnam in January 2019. Brunei, Chile, Malaysia, and Peru have signed the agreement but have yet to ratify it. A consolidated version of the CPTPP is available at *Materials*, MEGAREG, https://www.iilj.org/megareg/materials/ [https://perma.cc/E3MA-HXJQ].

197. Thomas Streinz, *Digital Megaregulation Continued: The Regulation of Cross-Border Data Flows in International Economic Law*, JAPAN SPOTLIGHT, July/Aug. 2020, at 49.

198. *See* Streinz, *supra* note 10, at 15–16.

Even in the absence of governmental limitations on data transfers, the internet exhibits multiple stratifications, hierarchies, and ultimately inequalities, contradicting the narrative that it is an inherently egalitarian infrastructure. International economic law is tilted against governmental distortions of what is perceived to be the natural and market-efficient transnational flow of factors of production, including data.[199] It is largely silent, however, about the hierarchies and power differentials that persist within the private sector. One such hierarchy concerns the relationship between different types of ISPs and their relationship to internet content providers.

The biggest internet backbone networks own operating infrastructure and interconnect with other networks of a similar size under only thinly (if at all) legalized "peering" arrangements.[200] Smaller ISPs, often regional, utilize a combination of paid transit under contractual terms and peering to deliver internet traffic.[201] Lower-tier ISPs depend on higher-tier ISPs and purchase access rights from them.[202] Internet Exchange Points ("IXPs") serve as the physical venue where ISPs can interconnect, thereby constituting central entities in an otherwise distributed network.[203]

All these arrangements are largely governed by private law technologies (especially contracting), if not mere social practice, within public law frameworks of telecommunications law.[204] Net

199. *See* Thomas Streinz, *International Economic Law's Regulation of Data as a Resource for the Artificial Intelligence Economy, in* ARTIFICIAL INTELLIGENCE AND INTERNATIONAL ECONOMIC LAW 175, 181 (Shin-yi Peng, Ching-Fu Lin & Thomas Streinz eds., 2021).

200. The basic idea behind "peering" is that computer networks mutually benefit from connecting with each other. Leonard Kleinrock, *Creating a Mathematical Theory of Computer Networks*, 50 OPERATIONS RSCH. 125, 128 (2002). Hence, in principle, there is no need for extensive bargaining (and contracting), at least not between large networks. For more on the late 20th century economics of peering (including between networks of different sizes), see generally, Pio Baake & Thorsten Wichmann, *On the Economics of Internet Peering*, NETNOMICS 89 (1999), which analyzes the economic rationales behind peering choices and Jean-Jacques Lafont, Scott Marcus, Patrick Rey & Jean Tirole, *Internet Peering*, 91 AM. ECON. REV. 287, 287–88 (2001).

201. *See generally* JAMES F. KUROSE & KEITH W. ROSS, COMPUTER NETWORKING: A TOP-DOWN APPROACH (8th ed. 2021) (discussing different "tiers" of networks).

202. *See* Baake & Wichmann, *supra* note 200, at 90, 92.

203. *See* Nikolaos Chatzis, Georgios Smaragdakis & Anja Feldmann, *On the Importance of Internet eXchange Points for Today's Internet Ecosystem*, ARXIV, at 2–3, https://arxiv.org/pdf/1307.5264v2.pdf [https://perma.cc/9GEL-UPKP].

204. *See generally* Uta Meier-Hahn, *Internet Interconnection: How the Economics of Convention Can Inform the Discourse on Internet Governance*, GIGANET: GLOBAL INTERNET

neutrality laws have become the most prominent intervention to guard against data flow discrimination on the internet.[205] Net neutrality laws prevent ISPs from leveraging the centrality of their role for internet access and traffic to charge their customers on either side (content providers and end users) more for delivering certain content faster. However, the scope of conventional net neutrality is limited, as it only concerns the "last mile" relationship between ISPs and end users, where ISPs are barred from deliberately favoring certain internet traffic.[206] The resulting market dynamics are complicated and subject to much debate.[207] For the purposes of this Article, we only seek to highlight that conventional net neutrality laws fail to account for disparate infrastructural control over data flows. Net neutrality laws solidify fundamental decisions enshrined in the internet's foundational protocols that were designed to deliver packets across inter-connected networks regardless of addressee or content.[208] While this may seem

---

GOVERNANCE ACADEMIC NETWORK, ANNUAL SYMPOSIUM 1 (2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2809867 [https://perma.cc/H7B7-V9WH] (analyzing internet interconnection arrangements).

205. Tim Wu is usually credited with inventing the term "net neutrality." *See generally* Tim Wu, *Network Neutrality, Broadband Discrimination*, 2 J. TELECOMM. & HIGH TECH. L. 141 (2003). *See also* BARBARA VAN SCHEWICK, INTERNET ARCHITECTURE AND INNOVATION 220 (2010).

206. In the United States, the Federal Communications Commission (FCC) issued the Open Internet Order in 2015, which classified Internet Service Providers as "common carriers" under Title II of the Communications Act of 1934 and Section 706 of the Telecommunications Act of 1996, thereby imposing net neutrality obligations on ISPs. 80 Fed. Reg. 19,738 (Apr. 13, 2015) (to be codified at 47 C.F.R. pts. 1, 8, 20). This decision was reversed in 2017. 33 FCC Rcd. 311 (2018).

In September 2018, California enacted its own net neutrality legislation, the California Internet Consumer Protection and Net Neutrality Act of 2018, which clarifies unlawful conduct by ISPs, including the controversial practice of zero rating (not charging for a certain kind of internet traffic). CAL. CIV. CODE § 3100 (West 2019). Net neutrality in the European Union is established through Regulation 2015/2120, laying down measures concerning open internet access. Regulation 2015/2120, 2015 O.J. (L 310) 1. The ECJ clarified that the regulation covered zero rating practices. Case C-807/18, Telenor Magyarország Zrt. v. Nemzeti Média- és Hírközlési Hatóság Elnöke, ECLI:EU:C:2020:708 (Sept. 15, 2020).

207. *See, e.g.*, Gerald R. Faulhaber, *Economics of Net Neutrality: A Review*, 3 COMMC'NS & CONVERGENCE REV. 53, 54 (2011). *See generally* Shane Greenstein, Martin Peitz & Tommaso Valletti, *Net Neutrality: A Fast Lane to Understanding the Trade-Offs*, 30 J. ECON. PERSPS. 127 (2016) (reviewing the existing literature on net neutrality).

208. For the basic idea behind the end-to-end principle, see generally J.H. Saltzer, D.O. Reed & D.D. Clark, *End-to-End Arguments in System Design,* 2 ACM TRANSACTIONS ON COMPUT. SYS. 4 (1984), which outlines the principles behind end-to-end system design. For the design implications, see generally David D. Clark, DESIGNING AN INTERNET (INFORMATION POLICY) (2018). *See* Nicholas P. Doty, Enacting Privacy in Internet Standards 17–23 (2020)

egalitarian, it also lends itself to unequal data flow dynamics. If all internet traffic is to be delivered equally to end users, those with much internet traffic stand to gain more than those with little.

As major platforms transformed the internet from a communication infrastructure into a data-gathering infrastructure, they found ways to leverage their infrastructural advantage over traditional ISPs (user access providers and data transit providers) to achieve preferential treatment for their data flows. The "last mile" treatment of internet traffic by ISPs is not the only factor that determines how quickly and reliably data is being transmitted on the internet. Platform companies invested heavily into their own content networks, which allowed them to bypass even traditional top-tier networks.[209] Sometimes ISPs entered into direct peering arrangements with platforms, leading to superior network performance without violating net neutrality laws.[210] Overall, platforms have increasingly relied on CDNs, especially to facilitate streaming services and cloud computing applications.[211] CDNs replicate content on servers that are physically or virtually closer to end users by maintaining a presence in, or close to, many large edge networks, thereby enhancing user experience.[212]

In sum, privileged access to fiber-optic submarine cables (especially in regions suffering lack of bandwidth), peering arrangements with ISPs, and CDNs with data centers at the edge of the networks enabled large platforms to deliver their content faster and more reliably than their competitors. These infrastructural dynamics remain largely unregulated by existing telecommunications law. Despite widespread adoption of net neutrality laws and protocols that,

---

(Ph.D. dissertation, U.C. Berkeley), https://npdoty.name/writing/enacting-privacy/drafts/enacting-privacy-20201219.pdf [https://perma.cc/E9TV-A8JF] (analyzing controversies in the W3C). *See generally* Michael Rogers & Grace Eden, *The Snowden Disclosures, Technical Standards, and the Making of Surveillance Infrastructures*, 11 INT'L J. COMMC'N 802 (2017) (scrutinizing the role of intelligence agencies in technical standard-setting organizations); Niels ten Oever, *"This Is Not How We Imagined It": Technological Affordances, Economic Drivers, and the Internet Architecture Imaginary*, 23 NEW MEDIA & SOC'Y 344 (2021) (describing the prioritization of corporate interests over the interests of end users in internet governance bodies).

209. *See* Philippa Gill et al., *The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse?*, *in* PASSIVE AND ACTIVE NETWORK MEASUREMENT 7–8 (Mark Claypool & Steve Uhlig eds., 2008).

210. *See* Sravan Patchala et al., *On the Economics of Network Interconnections and Its Impact on Net Neutrality*, 18 IEEE TRANSACTIONS ON NETWORK & SERV. MGMT. 4789, 4789 (2021).

211. *What is a CDN?*, CLOUDFLARE, https://www.cloudflare.com/learning/cdn/what-is-a-cdn [https://perma.cc/AV8K-J44A].

212. *Id*. at 4790, 4797–98.

in principle, do not discriminate between internet traffic, data does not "flow" equally on the internet.

B. Data Ownership

While the inequality of data flows tends to be underappreciated, the unequal distribution of data and the associated possibilities for value generation are well established.[213] Some have suggested that property law may rectify this situation by recognizing data ownership rights and by facilitating more efficient data markets.[214] However, as we have argued above, concentrated control over data is often a function of concentrated control over data infrastructures. Most data accumulation occurs irrespective of legal property rights in data, and technological means can be deployed to prevent data access by others, thereby entrenching data inequality in a way that is not dependent on property law.[215]

This does not mean, however, that the law of data ownership is irrelevant.[216] Certain categories of data are protected by established intellectual property (IP) rights: namely copyright, trade secrecy, and sui generis database rights recognized in some jurisdictions, such as the European Union's 1996 Database Directive.[217] Assertions of property claims are often invoked and can become contentious in response to demands for transparency, calls to share data with broader constituencies, and when mandatory data sharing is contemplated. Refusing to establish or recognize legal ownership rights in data is insufficient to address data inequality because unequal control over data can be asserted infrastructurally. At the same time, proactively establishing or recognizing legal property rights in data can *further*

213. *See, e.g.*, DAN CIURIAK, ECONOMIC RENTS AND THE CONTOURS OF CONFLICT IN THE DATA-DRIVEN ECONOMY 8 (2020); ERIC A. POSNER & E. GLEN WEYL, RADICAL MARKETS ch. 5 (2018).

214. *See, e.g.*, Jeffrey Ritter & Anna Mayer, *Regulating Data as Property: A New Construct for Moving Forward,* 16 DUKE L. & TECH. REV. 220, 221 (2018). *See generally* NÉSTOR DUCH-BROWN ET AL., THE ECONOMICS OF OWNERSHIP, ACCESS AND TRADE IN DIGITAL DATA (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2914144 [https://perma.cc/R2S8-KQCU] (analyzing non-rival data with property law concepts).

215. *See* discussion *supra* Section I.C.

216. *See* TERESA SCASSA, DATA OWNERSHIP 2 tbl.1 (2018); *see also* EURO. COMM'N STUDY ON EMERGING ISSUES OF DATA OWNERSHIP, INTEROPERABILITY, (RE-) USABILITY AND ACCESS TO DATA, AND LIABILITY 15 (2018) (prepared for the European Commission's DG CONNECT).

217. Council Directive 96/9/EC of Mar. 11, 1996, On the Legal Protection of Databases, 1996 O.J. (L 77) 20 [hereinafter Database Directive].

entrench infrastructural control with the authority of law by preventing redistributive measures, as data holders would then use property rights as an additional shield to exclude others from access.

International IP law and international investment law may shape the evolving and contested law of data ownership.[218] As international IP law became part of international trade and investment law, its conceptualization shifted from creating incentives to commodification (i.e., enabling trade) and assetization (i.e., protecting investments).[219] Data may be on a similar trajectory.[220]

In light of considerable legal uncertainty around data ownership, some have suggested establishing new ownership rights over data for "data creators" to facilitate contracting over data and to incentivize data generation.[221] This idea, however, ignores the not-IP-like incentive structure under which most data gets generated and rewards those who have treated data essentially as a *res nullius*: "things that belong to no one but can be claimed by whoever catches them first."[222] The comparison with established IP rights over data is instructive to validating this critique, as it clarifies the limited extent to which data is protected as property under existing IP law, while also raising the question of whether IP law indeed is the right legal framework for a discussion of data generation and its distributive effects.[223] The reason why IP law has framed this debate to date is likely due to path dependencies arising from data being intangible, as certain intangibles are already subject to IP protection. The discourse is often plagued by conflating the normative case for recognizing property rights in personal data to address concerns around individual privacy with broader questions about whether data, both personal and non-personal, already lends itself to property protections under existing law. For this reason, we first discuss the salience of "data

218. *See* Streinz, *supra* note 199, at 177.

219. *See* Rochelle Dreyfuss & Susy Frankel, *From Incentive to Commodity to Asset: How International Law Is Reconceptualizing Intellectual Property*, 36 MICH. J. INT'L L. 557, 560–75, 601–02 (2015).

220. *See infra* notes 291–294 and accompanying text.

221. *See, e.g.*, Ritter & Mayer, *supra* note 214, at 223, 260–76.

222. Pistor, *Rule by Data: The End of Markets?*, *supra* note 137, at 107.

223. *See* Rochelle C. Dreyfuss, *The Challenges Facing IP Systems: Researching for the Future*, 4 KRITIKA: ESSAYS ON INTELLECTUAL PROPERTY 1, 3 (2020):

> Are the changes to the creative environment so extensive that the terms on which traditional IP law operates are no longer functioning effectively? Are the piecemeal legal responses seen to date a first-best solution or are there better ways for the law to support, manage, and structure innovation in this new Age? Are the right parties profiting? What are the distributive effects of these changes and have they been properly taken into account?

ownership" under (domestic and international) IP law, as well as contract and tort law, before exploring interventions that adopt a property framing to mitigate data inequality.

Copyright law, which is internationally harmonized through the Berne Convention and the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement), protects the original *expression* of an idea in creative works, including in digital form, but not ideas, procedures, methods of operation, or mathematical concepts as such.[224] Traditional copyright, initially developed for the creation of artistic and literary works by individuals (e.g., authors and other creators), is an imperfect fit for data generation. A photograph, for example, captures reality (in some way). Yet, copyright does not attach to the facts contained within the picture, but merely protects the way in which these facts are being represented (i.e., expressed). Indeed, as the U.S. Supreme Court has stated, "facts—scientific, historical, biographical, and news of the day, may not be copyrighted and are part of the public domain available to every person."[225] Whether data thus constitutes fact becomes a critical question.[226] The

224. Berne Convention for the Protection of Literary and Artistic Works (as amended on Sept. 28, 1979) art. 2, Nov. 18, 1984, 102 Stat. 2853, 1161 U.N.T.S. 3; Agreement on Trade-Related Aspects of Intellectual Property Rights art. 9(2), Apr. 15, 1994, 108 Stat. 4809, 1869 U.N.T.S. 299 [hereinafter TRIPS]. *See generally* Michael Lehmann, *TRIPs, the Berne Convention, and Legal Hybrids*, 94 COLUM. L. REV. 2621 (1994) (arguing against a new IP paradigm for computer software).

225. Feist Publ's, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 348 (1991;); *see also* CH Canadian Ltd. v. L. Soc'y' of Upper Can.,. [2004] 1 S.C.R. 339, 352 ¶ 15 (Can.) ("[C]opyright protection only extends to the expression of ideas as opposed to the underlying ideas or facts.").

226. *See e.g.*, New York Mercantile Exch., Inc. v. IntercontinentalExchange, Inc., 497 F.3d 109, 114 (2d Cir. 2007) (in deciding whether settlement prices generated by an algorithm created by the plaintiff were subject to copyright protection, the court focused on whether the plaintiff was the *author* of the settlement prices or merely their *discoverer*); *see also* RBC Nice Bearings, Inc. v. Peer Bearing Co., 676 F. Supp. 2d 9, 21–23 (D. Conn. 2009) (data derived from a series of calculations carried out by the plaintiffs was unprotectable facts); BanxCorp v. Costco Wholesale Corp., 978 F. Supp. 2d 280, 300 (S.D.N.Y. 2013) (focusing on whether the formula used to convert raw data into the final value has "the degree of consensus and objectivity" that renders the final value "fundamentally a 'fact'"). These cases thus suggest that, in some circumstances, producers of indicators and other extrapolations may claim ownership, and enjoy copyright protection, in the indicators themselves. However, as Teresa Scassa points out:

> [T]o the extent that [such derived data] represent the idea behind the analytics that led to their creation [and thus] reflect a merger of idea and expression . . . , then it would seem that derived data must necessarily remain in the public domain, except where there is no merger between idea and expression. The challenge will be in determining when no merger occurs.

SCASSA, *supra* note 216, at 9.

TRIPS Agreement clarified that "compilations of data, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creation shall be protected *as such*."[227] Such protection, however, does not extend to *the data itself* and is without prejudice to any copyright subsisting in the data itself.[228] Hence, certain categories of data can be subject to copyright—if the general standard for creative works is satisfied—and compilations of data (databases) can be subject to copyright, too, if they constitute intellectual creations.[229] However, in reality, much data and most databases do not fulfill these requirements, as much data generation consists of recording facts, and most databases do not satisfy the threshold for creative works.[230]

The European Union responded to this (perceived) problem by creating a sui generis right for databases through its 1996 Database Directive.[231] Even though the European Union tried to entice other

227. TRIPS, *supra* note 224, art. 10(2) (emphasis added).

228. *Id.*

229. The Copyright Act, 17 U.S.C. §§ 101–810, for example, defines a compilation as a "collection and assembling of preexisting materials or of data that are selected in such a way that the resulting work as a whole constitutes an original work of authorship." *Id*. § 101. It also clarifies, in line with TRIPS, that the copyright in a compilation extends only to the compilation itself, and not to the underlying data. *Id*. § 103(b). Thus, for example, a compiler of genetic sequence data can ensure that her database is copyrightable if she chooses an original set of genes or proteins for inclusion in the database or arranges the database in an original manner. *See* M. Scott McBride, *Bioinformatics and Intellectual Property Protection*, 17 BERKELEY TECH. L.J. 1331, 1348–49 (2002). The copyright protection afforded to a compilation of facts, however, only applies to the elements that are deemed sufficiently creative or original. Therefore, in the case of a genetic sequence database deemed sufficiently creative/original under the *Feist* standard, 499 U.S. at 348, to merit copyright, the protection afforded by that copyright would extend only to the compiler's original selections or arrangement of data, not the data as such. McBride, *supra*, at 1349.

230. As Teresa Scassa observes:

> [C]ompilations of fact present many challenges when it comes to copyright. 'Whole universe' sets of fact may not reflect an original selection; similarly, where facts are arranged according to industry norms or standards, the compilation may lack originality. A dataset that is constantly growing (for example, streamed sensor data) may similarly be incapable of being a compilation since there is never a completed work. Even if a selection or arrangement is original, the principle that facts are in the public domain means that only the original selection or arrangement of the compilation will be protected; anyone who extracts facts from the compilation using an independent selection and arrangement of those facts has not infringed copyright.

SCASSA, *supra* note 216, at 6.

231. Database Directive, *supra* note 217, ch. 3.

jurisdictions to reciprocate,[232] most jurisdictions, including Canada and the United States, have refrained from expanding copyright protection in this way.[233] In deviating from the creative works standard traditionally deployed in copyright law, the sui generis right applies when a database maker can show that "there has been qualitatively and/or quantitatively a *substantial investment* in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database."[234] The European Union's own review of its Database Directive found no evidence that the creation of a sui generis right had any impact on the production of databases.[235] This striking finding indicates that IP incentives are neither necessary nor sufficient to explain database creation or indeed data generation more broadly.[236] The fact that the European Union refrained from reforming the Database Directive by revoking sui generis data protection rights also illustrates how difficult it is to revoke property protections once established, even if they do not achieve their intended objective, as

232. *Id*. art. 11(3). The Database Directive foresaw international agreements between the EU and third countries, under which the EU's sui generis database protection would be extended to databases created in third countries, if those countries instituted sui generis database protection domestically. *Id.*

233. *Database Legal Protection*, FORSGREN FISHER MCCALMONT DEMAREAA TYSVER LLP, https://www.bitlaw.com/copyright/database.html [https://perma.cc/6PLD-JJCR]; McCarthy Tétrault LLP, *Big Data: Legal Aspects from a Canadian Perspective-Part I,* LEXOLOGY (Feb. 22, 2016), https://www.lexology.com/library/detail.aspx?g=1588fe0d-64fd-4b84-b3b4-ef75efc72d08 [https://perma.cc/D8EH-CRJD]. "Sui generis protection" means the protection that is afforded to a particular object by virtue of that object's unique characteristics. *Sui Generis*, BLACK'S LAW DICTIONARY (11th ed. 2019):

> The term [sui generis] is used in intellectual-property law to describe a regime designed to protect rights that fall outside the traditional patent, trademark, copyright, and trade-secret doctrines. For example, a database may not be protected by copyright law if its content is not original, but it could be protected by a sui generis statute designed for that purpose.

234. Database Directive, *supra* note 217, art. 7 (emphasis added). The European Court of Justice clarified that the investment must pertain to the creation of the database, not the data contained therein. Case C-46/02, Fixtures Marketing, 2004 E.C.R. I-10365 ¶ 33.

235. European Commission Staff, *Evaluation of Directive 96/9/EC on the Legal Protection of Databases of Apr. 25, 2018*, at 15–19, SWD (2018) 146 final.

236. This phenomenon is not necessarily confined to data generation. *See generally* Rochelle Cooper Dreyfuss, *Does IP Need IP? Accommodating Intellectual Production Outside the Intellectual Property Paradigm*, 31 CARDOZO L. REV. 1437 (2010).

beneficiaries can defend their established rights under international investment and even human rights law.[237]

Protection of undisclosed information, as required by the TRIPS Agreement and further ratcheted up in recent preferential trade agreements, operates under a different IP logic than copyright and lends itself to potentially vast protection of data as trade secrets. To acquire protection, the information in question must be secret in the sense that it is not generally known or accessible, must have commercial value due to being secret, and the entity lawfully in control of the information must have made reasonable steps to keep it secret.[238] The value proposition under which "big data" operates lends credence to the claim that (potentially any) data has commercial value. Generating, storing, and processing data through cloud computing infrastructures makes it easier to limit access to data and to keep it secret.[239] In other words, control over data infrastructures that generate and store data of commercial value in secure fashion is being rewarded by trade secrecy protection of such data, thereby solidifying data inequality and creating questionable incentives for competition and innovation.[240] Unlike copyright and patent law, which afford protection for only a limited period of time, trade secrecy affords potentially unlimited protection.[241] One limiting principle is that trade secrecy protection ends once the information in question is no longer secret.[242] Unlike patent law, which requires public disclosure and thereby facilitates the dissemination of knowledge (while retaining exclusive rights of commercial exploitation for a particular period of time), trade secrecy law incentivizes data holders to keep information secret to enable exclusive commercial exploitation (potentially

237. *See generally* Martin Husovec, *The Fundamental Right to Property and the Protection of Investment: How Difficult Is It to Repeal New Intellectual Property Rights*, *in* RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND INVESTMENT LAW (Christophe Geiger ed., 2019) (analyzing limitations arising from international investment law and the European Convention of Human Rights). On the human right of property, see generally Jose E. Alvarez, *The Human Right of Property,* 72 U. MIAMI L. REV. 580 (2018) (canvassing the existence of a human right to property).

238. TRIPS, *supra* note 224, art. 39.

239. Jeanne Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 N.Y.U. L. REV. 706, 718–20 (2019)

240. David S. Almeling, *Seven Reasons Why Trade Secrets Are Increasingly Important,* 27 BERKELEY TECH. L.J. 1091, 1092–95 (2012). *See generally* Fromer, *supra* note 239 (discussing technological changes, how that has made trade secrecy overly protective, and the costs that secrecy imposes).

241. Fromer, *supra* note 239, at 728–29.

242. Mark Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 STAN. L. REV. 311, 352 (2008).

forever).[243] Data holders can try to retain secrecy qua contractual arrangements (e.g., nondisclosure agreements), but they run the risk of losing trade secrecy protection if such safeguards fail. Trade secrecy is hence not an IP right that is suitable to the kind of market-driven data transactions envisioned by proponents of new property rights for data creators.[244] Instead, the significance of trade secrecy lies mainly in fending off attempts to get access to data, as in the case of attempts to scrutinize datasets which are suspected of contributing to bias in machine-learning.[245]

Property protections of data under IP law are not comprehensive, but data holders may still use contracts as legal technologies for regulating access to and control over data. In data contracts, data providers often assert "data ownership" even where arguably no recognized property right in data exists outside the contractual arrangement.[246] The language in "data licensing agreements" is often borrowed from or at least influenced by IP law (especially copyright), even when "data ownership" claims are tenuous.[247] The terms of such license agreements vary.[248] In the absence of explicit property protections, data providers and acquirers run the risk of third parties not bound by the contracts eventually acquiring the data. This creates an incentive to resort to technological means to control access to data at a distance. In other words, where legal control is insufficient or contested, infrastructural control might still suffice. Digital Rights Management (DRM) was initially developed for copyright protection; it then achieved significant

243. *See id.* at 352–53 (comparing trade secrecy to other IP rights and arguing that trade secrets should "expire" after a certain period).

244. *See generally* David S. Levine, *Secrecy and Unaccountability: Trade Secrets in our Public Infrastructure*, 59 FLA. L. REV. 135 (2007) (discussing the negative impact of trade secrecy on public infrastructure).

245. *See generally, e.g.*, Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018) (discussing the use of trade secret law to shield information in criminal prosecution).

246. *See, e.g.*, Daniel J. Gervais, *The Protection of Databases*, 82 CHI.-KENT L. REV. 1109, 1148 (2007).

247. *Id*.

248. *See generally* Daniel Glazer et al., *Data as IP and Data License Agreements*, Practical Law Practice Note 4-532-4243, THOMSON REUTERS (2019) (discussing provisions delineating ownership and use rights, including use restrictions, purpose limitations, stipulations regarding the (in)ability of the licensee to aggregate or modify the data or create or use other derivative data, treatment of derived data, provisions regarding data delivery, security obligations, audit rights, risk allocations (via warranties and indemnities), dispute resolution provisions, and other provisions standard to contracts (such as termination, assignment, choice of law, etc.)).

support from content industries and obtained legal protections against circumvention in many jurisdictions.[249] However, even without copyright and anti-circumvention protection of data by law, DRM technologies can still be deployed to control access to data.

Infrastructural analysis illuminates the ways in which legal technologies constitute, enable, and otherwise intersect with control over data.[250] There is no need for full property protection of each component or layer of data infrastructures to sustain data inequality as long as control over key components of the infrastructure as a whole is not being challenged. Even where legal ownership over data is in question, control over data can be exercised through control over data infrastructures. Conversely, even when data rights and obligations are allocated in contracts, effective monitoring and enforcement is often equally dependent on infrastructural control (e.g., by allowing only access on certain machines).

Data infrastructures have both physical and digital components. The law of property, as it relates to real and personal (both tangible and intangible) property, protects physical components of data infrastructures. Laws regarding the possession, use, and control of real and personal property allow for the ownership of the physical components of data infrastructures (e.g., cables, cell towers, data centers, and computers) and protect owners' right to decide who has access to these objects (e.g., through the law of trespass). Data is always stored somewhere, though not necessarily in one place,[251] and the prerequisite hardware (e.g., hard drives) enjoys property protection (e.g., against theft or destruction), as does the real property on which data centers reside. In many jurisdictions, computer systems enjoy additional protections against unauthorized access, which can lead to claims functionally equivalent to the right to exclude others from

249. *See* Digital Millennium Copyright Act, 17 U.S.C. § 1201 (prohibiting circumvention of copyright protection systems). For a sharp critique of DRM from a development perspective, see *Digital Rights Management (DRM): A Failure in the Developed World, A Danger to the Developing World*, *ITU-R Working Party 6M Report on Content Protection Technologies,* ELEC. FRONTIERS FOUND. (Mar. 11, 2005), https://www.eff.org/wp/digital-rights-management-failure-developed-world-danger-developing-world [https://perma.cc/AL5E-ZZWU].

250. The distinction between data as an object or resource and the data infrastructures that produce, shape, store, and transfer data is helpful in clarifying where ownership rights exist and where such legal protection is non-existent or at least contested. *See supra* Section I.A; *infra* notes 251–262 and accompanying text.

251. *See* Paul M. Schwartz, *Legal Access to the Global Cloud*, 118 COLUM. L. REV. 1681, 1689–99 (2018) (discussing different models of cloud computing and the implications of "data sharding" for governmental access to data stored in transnational cloud infrastructures).

accessing data on the basis of a property right.[252] For example, the Computer Fraud and Abuse Act (CFAA) prohibits intentionally accessing a computer without authorization.[253] LinkedIn claimed that web scraping and use of publicly available information in violation of platform terms are "without authorization" under the CFAA.[254] Tort law may provide a similar protection under the common law tort theory of "trespass to chattels." Courts in the United States have held that online providers of information can protect their databases from unauthorized use and copying under this theory.[255] The protection that this theory appears to offer only relates to the data infrastructure used to store or publish information, not to the data as such.[256]

Ownership over the *digital* components of data infrastructures varies. Those who exercise corporate control over data infrastructures do not necessarily own every component at every layer. Copyright law can be used to assert control over software, encompassing both source and object code, as well as structure, sequence, organization, and features generated by code.[257] Certain companies leverage their

252. *See* Pistor, *Rule by Data*, *supra* note 137, at 109 (arguing that both the CFAA and the European Union's Database Directive grant property-like legal protections data harvesters).

253. 18 U.S.C. § 1030(a)(2)(C).

254. hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 999 (9th Cir. 2019). The U.S. Supreme Court vacated and remanded the case in light of its decision in Van Buren v. United States, 141 S. Ct. 1648 (2021). LinkedIn Corp. v. hiQ Labs, Inc., 141 S. Ct. 2752 (2021). On cyber-trespass laws as quasi-property regimes, see generally Thomas E. Kadri, *Platforms as Blackacres*, 68 UCLA L. REV. 1184 (2022).

255. *See* Charles C. Huse, *Database Protection in Theory and Practice: Three Recent Cases*, 20 BERKELEY TECH. L.J. 23, 29 (2005); Margaret Jane Radin, *A Comment on Information Propertization and Its Legal Milieu*, 54 CLEV. ST. L. REV. 23, 25 (2006). According to this theory, "trespass to chattels" occurs when "an intentional interference with the possession of personal property has proximately caused injury." *See* Thrifty-Tel v. Bezenek, 54 Cal. Rptr. 2d 468, 473 (Ct. App. 1996); eBay v. Bidder's Edge, 100 F. Supp. 2d 1058, 1069–70 (N.D. Cal. 2000).

256. The party invoking the theory must be the owner of the "chattel" that has been interfered with, thus reverting to the question of what type of legal regime could confer rightful ownership interests over data as such. *See eBay*, 100 F. Supp. 2d at 1070.

257. TRIPS, *supra* note 224, art. 10(1) provides that computer programs, whether in source or object code, shall be protected as literary works under the Berne Convention of 1971. This is reflected in the copyright laws of jurisdictions around the world, e.g., Copyright Act, R.S.C. 1985, c. C-42, s. 5(1) (Can.); Code de propriété intellectuelle [Intellectual Property Code] art. L112-1 (Fr.); Gesetz über Urheberrecht und verwandte Schutzrechte [Urheberrechtsgesetz] [Act on Copyright and Related Rights], Nov. 28, 2018, Gesetsez [BGBI] § 2(1), 2014, as amended (Ger.); Copyright, Designs and Patents Act 1988, c. 48, ch. 1 § 3(1) (U.K.); and Copyright Act, 17 U.S.C. § 102 (2012). In the United States, if one wants

proprietary control over data management software and data formats to retain control over data. Farming data platforms in digital agriculture, introduced above,[258] are a case in point. Both Monsanto and John Deere use proprietary software which locks-in farmers and "their" data due to a lack of data portability and interoperability.[259] This digital vise wields distributive effects, as larger farms may be able to internalize the rising costs associated with dependencies on corporate data infrastructures better than their smaller competitors, thus further empowering "supersized farms" while marginalizing smaller farms, farmers, and their workers.[260] To the (limited) extent to which software enjoys patent protection in some jurisdictions, patent law may have comparable effects, albeit potentially counterbalanced by reasonable and non-discriminatory licensing terms.[261] While software is generally not patentable in most jurisdictions, the increasing integration of software into machines (e.g., robots) complicates the separation of software from hardware

---

to obtain copyright registration, source code for a specific version of the computer program must be deposited as part of the copyright registration process. Source code can be uploaded to the electronic registration system, as a PDF file, or as any other file type accepted by the U.S. Copyright Office. Alternatively, the source code can be printed out on paper and mailed to the Office. *Cf.* U.S. COPYRIGHT OFF., CIRCULAR 61: COPYRIGHT REGISTRATION OF COMPUTER PROGRAMS (2021).

258. *See* discussion *supra* Section I.D.

259. Rotz et al., *supra* note 129, at 117. As one of the farmers interviewed by Rotz and her colleagues observed:

> Everything is connected to the internet, I don't think you have any control over it anymore. That is a tricky one right, because like I said they [data management companies] have access to everything, yet we still get the bills all the time. So when do we get to issue a bill and get a little bit of a kick-back for the information that we are generating on a daily basis? Because, the supplier companies are like 'we need to find our R&D programs, to make it better for you guys.' But every time you make a new investment then the price of your equipment just went up because now it is the newest, latest, greatest, so you figure you [the company] can charge another 10% or another 5% or whatever amount it might be. So, you [the company] took all my information to do that.

Id.

260. For broader effects of digital farming, see, for example, Evagelos D. Lioutas et al., *Key Questions on the Use of Big Data in Farming: An Activity Theory Approach*, WAGENINGEN J. LIFE SCIS., Dec. 2019.

261. *See generally* Mark Lemley, *Software Patents and the Return of Functional Claiming*, 2013 WISCONSIN L. REV. 905 (2013) (arguing that courts should scrutinize overbroad functional claiming); Yann Meniere and Nikolaus Thumm, *Fair, Reasonable and Non-Discriminatory (FRAND) Licensing Terms*, JRC SCI. & POL'Y REP. 27333 (2015) (discussing licensing in the ICT sector).

and may lend patent protection to complex data infrastructures in which software and hardware are inseparably intertwined.[262]

Proprietary control over software is no longer the norm. Open source software, where copyright law is leveraged to make software freely available, has become increasingly important, acquiring infrastructural importance for a variety of use cases.[263] Examples range from encrypted data transfers (e.g., OpenSSL) over operating systems of web servers (e.g., Apache), to machine-learning algorithms (e.g., TensorFlow) and the internal infrastructures of large tech companies (e.g., Google runs a version of the open source operating system Linux on its desktop computers and servers.).[264] Corporate control over data infrastructures is commensurate with certain components of those infrastructures being available to everyone else.

In the United States, the question of whether copyright protection extended to certain application program interfaces (APIs) has been highly contested.[265] APIs are an important infrastructure for software, as they facilitate the interaction and interoperability between different components within and across programs and systems.[266] APIs can also be used for data transfers online, where web-based APIs ("webAPIs") have emerged as important vehicles for access to data.[267] Copyrightability of APIs is largely a question of rent-seeking and -distribution. Infrastructural control over APIs exists irrespective of

---

262. *See Hardware and Software*, EUR. PATENT OFFICE. (May 20, 2019), https://www.epo.org/news-events/in-focus/ict/hardware-and-software.html [https://perma.cc/8HQM-TSYL] (discussing patent claims for "computer-implemented inventions").

263. *Open Source Software as Digital Infrastructure: Legal Technologies and Institutional Design*, N.Y.U. SCH. OF L. GUARINI INST. FOR GLOB. LEGAL STUD., https://www.guariniglobal.org/events-overview/2020/1/14/open-source-software-as-digital-infrastructures-legal-technologies-amp-institutional-design [https://perma.cc/68L6-TNWZ].

264. Steven J. Vaughan-Nichols, *The Truth About Goobuntu: Google's In-House Desktop Ubuntu Linux*, ZDNET (Aug. 29, 2012, 11:12 PM), https://www.zdnet.com/article/the-truth-about-goobuntu-googles-in-house-desktop-ubuntu-linux/ [https://perma.cc/5R3T-3PA9]; Steven J. Vaughan-Nichols, *Google Moves to Debian for In-House Linux Desktop*, ZDNET (Jan. 18, 2018, 6:03 PM), https://www.zdnet.com/article/google-moves-to-debian-for-in-house-linux-desktop/ [https://perma.cc/R2YZ-DPZK].

265. *See generally* Peter S. Menell, *Rise of the API Copyright Dead? An Updated Epitaph for Copyright Protection of Network and Functional Features of Computer Software*, 31 HARV. J.L. & TECH. 305 (2018); Jonathan Band, *The Global API Copyright Conflict*, 31 HARV. J.L. & TECH. 615 (2018).

266. Menell, *supra* note 265, at 474.

267. Oscar Borgogno & Giuseppe Colangelo, *Data Sharing and Interoperability: Fostering Innovation and Competition Through APIs*, 35 COMPUT. L. & SEC. REV. no. 105314, 2019, at 3.

copyright protection.[268] The designers and operators of webAPIs decide who can access the dataset they make available in this way and under what terms (e.g., download, just view, and perform queries).[269] This is of particular concern when the resultant gatekeeping role can be abused to shut down access to data or to discriminate between different users. For related reasons, webAPIs are sometimes seen as inferior vehicles for effectuating rights to data portability, compared to user download and upload.[270] As webAPIs assume an increasing role as gateways between data infrastructures, those who control access to data qua webAPIs assume an infrastructural role regardless of copyright protection. For example, Ushahidi, a non-profit software company based in Nairobi, Kenya, provides its data collection, visualization, and interactive mapping services based on APIs provided by Google and Twitter.[271] If these companies decided to terminate or alter the API, Ushahidi's ability to continue its service would be negatively impacted.

Can unequal control over data be mitigated through legal instruments that stem from and operate within a property law framework? In the following paragraphs, we discuss such interventions, including "open data" in public and private sector contexts, mandatory data sharing, and the granting of property rights over personal data.

Promoters of "open data" seek to make data freely available.[272] However, unlike open-source software licensing, open data licensing cannot operate under the assumption that data is subject to copyright and has to account for the disparate sui generis IP protection of

268. Such control can be challenged by demanding interoperability. *See* IAN BROWN, INTEROPERABILITY AS A TOOL FOR COMPETITION REGULATION 1, 13 (2020), https://osf.io/preprints/lawarxiv/fbvxd/ [https://perma.cc/7NLD-NQ3T].

269. Public webAPIs tend to vest decision-making powers over the terms of data access solely with the data host. In private transactions, APIs are sometimes used alongside contracts to regulate access to data via legal and technical means.

270. GABRIEL NICHOLAS & MICHAEL WEINBERG, DATA PORTABILITY AND PLATFORM COMPETITION: IS USER DATA EXPORTED FROM FACEBOOK ACTUALLY USEFUL TO COMPETITORS? 2, 21 (2019), https://www.law.nyu.edu/sites/default/files/Data%20Portability%20and%20Platform%20Competition%20-%20Is%20User%20Data%20Exported%20From%20Facebook%20Actually%20Useful%20to%20Competitors.pdf [https://perma.cc/38BZ-N3E8].

271. *About Ushahidi*, USHAHIDI, https://www.ushahidi.com/about [https://perma.cc/TJ5H-57CS].

272. *See, e.g.*, Beth Simone Noveck, *Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency*, 19 YALE HUM. RTS. & DEV. L.J. 1 (2017) (highlighting the advantages of open government data over Freedom of Information Act (FOIA) requests).

databases across jurisdictions.[273] While modern "open data" licenses can account for these variations (at the expense of being more complicated and cumbersome),[274] the unwillingness of data hoarders to make data available as "open data" may be more difficult to tackle. In reality, with some exceptions,[275] "open data" strategies have largely been deployed to make certain categories of data generated by the public-sector data available as "open data." Recent instruments of international economic law actively encourage this approach.[276] Corporate data infrastructures often serve as gateways to find "open" data sets or include "open data" in their cloud offerings.[277]

Notably, open data licenses do not guard against de facto appropriation and exploitation of "open data" by those with concentrated control over data infrastructures. The unrestricted availability of "open data" can lead to questionable distributive outcomes, especially when this model is only being deployed to make public sector data available to the private sector, but not vice versa. Despite the plethora of reports, articles, and documents professing the value of "open data,"[278] there do not seem to be rigorous studies comparing the value derived from open data between public and commercial actors, although a few scholars have suggested that open data disproportionately empowers commercial entities already holding

273. *See supra* notes 247–249 and accompanying text.

274. Examples include the Open Database License ("ODbL") by Open Data Commons, which belongs to Open Knowledge International. The Creative Commons ("CC") licensing system, initially developed for creative works, has been adjusted to accommodate use for data. *See generally* Alexandra Giannopoulou, *Understanding Open Data Regulation: An Analysis of the Licensing Landscape*, *in* OPEN DATA EXPOSED 101 (Bastiaan van Loenen, Glenn Vancauwenberghe & Joep Crompvoets eds., 2018).

275. In 2016, Google released Open Images, a dataset of more than 9 million images labelled according to over 6000 categories. Ivan Krasin & Tom Duerig, *Introducing the Open Images Dataset*, GOOGLE BLOG (Sept. 30, 2016), https://ai.googleblog.com/2016/09/introducing-open-images-dataset.html [https://perma.cc/R9JL-C7QS].

276. *See, e.g.*, Protocol Replacing the North American Free Trade Agreement with the Agreement Between Canada, the United States of America, and the United Mexican States, Can.-U.S.-Mex. art. 19.18, Nov. 30, 2018, 107 Stat. 2116 [hereinafter USMCA].

277. *See, e.g.*, *Dataset Search*, GOOGLE, https://datasetsearch.research.google.com [https://perma.cc/V5LR-LHXE]; *Registry of Open Data on AWS*, AMAZON WEB SERVS., https://registry.opendata.aws [https://perma.cc/39VQ-YBXU].

278. *See* Stefaan Verhulst et al., *Open Data Index: 10 Insights on the Value of Open Data*, MEDIUM (Apr. 21, 2020), https://medium.com/open-data-policy-lab/open-data-index-10-insights-on-the-value-of-open-data-f810e7cb8e9 [https://perma.cc/ZH4F-683V] (referencing the Open Governance Research Exchange database, *Open Data*, OPEN GOVERNANCE RSCH. EXCH., https://ogrx.org/categories/open-data.html [https://perma.cc/J7YB-88WN], which features hundreds of studies).

a socially and economically advantageous position.[279] Open data arrangements allow companies to access data they do not yet control at zero cost. In this way, public data generation is being repurposed to subsidize highly profitable and data-rich companies. For example, recent advances in machine translation can be attributed, in part, to human translation work that employees at international organizations, such as the United Nations and the European Union, have carried out for decades, as companies in the machine learning business can exploit these documents to train their machine-learning algorithms.[280]

In contrast to voluntary "open data" initiatives, forcing private data holders to give up data by way of mandatory data sharing poses a direct legal challenge to de facto control over data. If faced with such requests, data holders are likely to resort to legal data ownership claims to counter others' purported data access rights. While access to data rights can lead to effective data redistribution, they also entrench the choices that data generators made in the process of datafication. Whether entities make data available as "open data" or are forced to share data under mandatory data sharing laws, many choices regarding the terms of data access (e.g., file formats) and the possibilities of data use (e.g., with regard to categorization and structuring of data) have already been made by data producers, who have the power and means to determine which data is collected in the first place and how. The decision of *what* data to produce and *which* data to grant access to often rests entirely with data-generating entities.[281] Discrepancies in data utility are further compounded if those who generate data also have

279. Michael B. Gurstein, *Open Data: Empowering the Empowered or Effective Data Use for Everyone?*, 16 FIRST MONDAY (2011), https://journals.uic.edu/ojs/index.php/fm/article/view/3316/2764 [https://perma.cc/LYX7-2KNF]; Bianca Wiley, *Open Data Endgame: Countering the Digital Consensus,* CIGI PAPERS NO. 186, at 5 (Aug. 2018), https://www.cigionline.org/publications/open-data-endgame-countering-digital-consensus [https://perma.cc/2QX3-JAJH].

280. *See* Ido Ramati & Amit Pinchevski, *Uniform Multilingualism: A Media Genealogy of Google Translate*, 20 NEW MEDIA & SOC'Y 2550, 2556 (2018) (analyzing the underlying power structure of algorithmic and human collaboration in Google translate):

> When Google Translate was launched in 2006, it began utilizing texts like United Nations documents, international treaties, and multilingual corporate websites, all of which were accessible through its various services: in the words of Translate's first architect Franz Josef Och, its algorithms started mining 'everything that's out there.'

On broader concerns about machine translations see, for example, Shlomit Yanisky-Ravid & Cynthia Martens, *From the Myth of Babel to Google Translate: Confronting Malicious Use of Artificial Intelligence—Copyright and Algorithmic Biases in Online Translation Systems*, 43 SEATTLE UNIV. L. REV. 99, 100 (2019).

281. *See* discussion *supra* Section I.C.

access to other data and the means to aggregate and process data from different sources.

Another idea that seeks to deploy property law to challenge corporations' factual control over data proposes to award new property rights over personal data to individuals. This idea is not new,[282] resurfaces regularly,[283] and straddles the domains of property and data protection law.[284] In our view, it is analytically helpful to distinguish between individual data ownership rights derived from property law and data rights stemming from data protection laws. Certain data protection rights (such as the right to erasure) can be conceptualized as akin to property rights (which commonly also include the right to destroy one's own property).[285] However, such comparisons risk obscuring the different logics animating the respective legal regimes.[286] At the same time, the idea to accord individual property rights over one's personal data shares some of the conceptual difficulties that also plague data protection law. One such problem concerns the distinction between personal and non-personal data: Data inequality also accrues due to the concentration of infrastructural control over non-personal data, but approaches that depend on individuals exercising *their* rights over *their* data cannot account for this. A related problem is that data is inherently relative, which makes it difficult to delineate rights and obligations.[287] The case of genetic data made available to genetic data analytics companies such as 23andMe is illustrative of this phenomenon: Does or should one "own" one's genetic data and have the right to make it available to others if this also reveals genetic information over one's relatives? Even if these conceptual difficulties could be overcome, an individualistic data property rights regime seems unlikely to be effective at challenging infrastructural control over data. The value proposition of "big data"

282. *See, e.g.*, Pamela Samuelson, *Privacy as Intellectual Property*, 52 STAN. L. REV. 1125, 1130 (1999).

283. *See, e.g.*, ERIC A. POSNER & E. GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY 205 (2018); Katrina Miriam Wyman, *Property in Radical Markets*, 87 U. CHI. L. REV. 125, 125–27 (2019).

284. *See generally* Ignacio Cofone, *Beyond Data Ownership*, 43 CARDOZO L. REV. 507 (2021) (rejecting property approaches to personal data and arguing for privacy law solutions instead).

285. *See, e.g.*, Jacob M. Victor, Comment, *The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy*, 123 YALE L.J. 513, 524–25 (2013).

286. *But see* Diana Liebenau, *What Intellectual Property Can Learn from Informational Privacy, and Vice Versa*, 30 HARV. J. L. & TECH. 285, 286–88 (2016).

287. Sebastian Benthall, *Situated Information Flow Theory*, PROCEEDINGS OF THE 6TH ANNUAL HOT TOPICS IN THE SCIENCE OF SECURITY §§ 1, 2 (HOTSOS) (2019); *see also* Salome Viljoen, *A Relational Theory for Data Governance*, YALE L.J. 573, 603–16 (2021).

entails that each individual's personal data is far less valuable compared to a dataset aggregating the data of many people, which complicates bargaining over adequate compensation for access to data. Moreover, the established power asymmetries and collective action problems that pit individual interests, even if protected by property claims, against concentrated corporate interests and infrastructural control would be difficult if not impossible to overcome.[288]

If control over data is largely a function of control over data infrastructures and not entirely dependent on recognizing legal ownership rights in data, then interventions that recognize such ownership rights are unlikely to be effective in remedying uneven control over data. In fact, such interventions might even produce adverse effects by rewarding those who already enjoy infrastructural control with additional legal safeguards against redistributive and other regulatory measures. Indeed, there is reason to think that a property framing is inapposite to questions of uneven control over data, precisely because it ignores the infrastructural dimensions of control over data and the related power to datafy. The most radical proposition along these lines would be to declare the world of data a *res communis*, a public good that cannot be owned by anyone, rather than a *res nullius* that is up for grabs.[289] The economic argument for this approach is a continuation of information economics, which has identified information asymmetries as harmful for economic growth and advocated for knowledge as a global public good, with only relatively thin IP protections.[290] While this approach has its own complications, as it would need to be reconciled with established IP and data protection rights, it strikes us as the right baseline to develop appropriate data ownership and control regimes going forward.

In the meantime, it seems likely that international IP and investment law will be used to frame this debate in terms of commodification and asset protection. As Rochelle Dreyfuss and Susy Frankel have explored in detail, the incorporation of the international IP regime into international trade law turned what was once developed

288. This problem also plagues the effectiveness of individual data rights that data subjects commonly enjoy under data protection laws. *See* discussion *infra* Section II.C.

289. *See* Pistor, *Rule by Data*, *supra* note 137, at 124 (arguing that states should have declared personal data *res communis*); *see also* Richard A. Epstein, *Property Rights and Governance Strategies: How Best to Deal with Land, Water, Intellectual Property, and Spectrum*, 14 COLO. TECH. L. J. 181, 182 (2016) (contrasting *res nullius* with *res communis*).

290. *See e.g.*, Joseph Stiglitz, *Economic Foundations of Intellectual Property Rights*, 57 DUKE L.J. 1693, 1714–15 (2008); DEAN BAKER, ARJUN JAYADEV & JOSEPH STIGLITZ, INNOVATION, INTELLECTUAL PROPERTY AND DEVELOPMENT: A BETTER SET OF APPROACHES FOR THE 21ST CENTURY 35–39 (July 2017) (discussing pathologies of international intellectual property rights regime.

to coordinate incentives for individual IP creation into a system to protect the transnational interests of those who trade IP as a commodity (including within firms as part of sophisticated tax avoidance strategies).[291] International investment law provides additional safeguards under which IP is being conceptualized as an asset to be protected against direct or indirect expropriation or regulation that can be challenged as a violation of the notoriously vague "fair and equitable treatment" standard.[292] As international investment law also shapes private law,[293] it risks imposing a data-as-a-resource framing onto the evolving debate around data ownership, and it is likely to be mobilized against attempts to redistribute data qua mandatory data sharing.[294]

C. Data Rights

The rights-based approach to data regulation in the form of data protection and privacy laws has dominated the discourse around legal regulation of the digital transformation of economies and societies around the globe. Data protection and data privacy laws emerged in the 1970s in response to advances in computation technology.[295] The Organisation for Economic Co-operation and Development (OECD)'s Privacy Guidelines of 1980 and the Council of Europe's Data Protection Convention of 1981 created two early models, respectively, to internationally harmonize privacy principles and data protection laws.[296] When the European Union harmonized its Member States'

291. *See generally* Dreyfuss & Frankel, *supra* note 219.

292. *See, e.g.*, Benedict Kingsbury & Stephan Schill, *Investor-State Arbitration as Governance: Fair and Equitable Treatment, Proportionality and the Emerging Global Administrative Law*, 8–18 (IILJ Working Paper 2009/6) (distilling different normative principles present in arbitral jurisprudence on fair and equitable treatment and discussing implications for the exercise of public power).

293. *See generally* Julian Arato, *The Private Law Critique of International Investment Law*, 113 AM. J. INT'L L. 1 (2019) (arguing that international investment law constrains how nations may organize the laws of property, contracts, corporations, and intellectual property).

294. *See* Thomas Streinz, *International Economic Law's Regulation of Data as a Resource for the Artificial Intelligence Economy*, *in* ARTIFICIAL INTELLIGENCE AND INTERNATIONAL ECONOMIC LAW: DISRUPTION, REGULATION, AND RECONFIGURATION 175, 187–88 (Shin-yi Peng, Ching-Fu Lin & Thomas Streinz eds., 2021).

295. *See* Przemysław Pałka, *Data Management Law for the 2020s: The Lost Origins and the New Needs*, 68 BUFF. L. REV. 559, 572–89 (2020).

296. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, OECD, https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

data protection laws through its 1995 Data Protection Directive,[297] it created a new template for data protection law, which exercised a significant compliance pull outside Europe.[298] The European Union's GDPR carried forward this legacy. The GDPR is routinely touted as a template for jurisdictions around the world that do not yet have data protection laws or are planning to reform their existing laws.[299] The GDPR dovetails with the Council of Europe's Reformed Convention 108, which is open for non-European countries to join.[300] By the end of 2019, 142 countries had data privacy laws on the books, sixty-two countries more than in the previous decade.[301] These are important developments, as the world ostensibly gravitates towards comprehensive data protection regulation.

One should be cautious, however, not to overestimate the extent of convergence, since differences persist even if data protection law on the books may look similar across jurisdictions. The European Union's data protection regime was reinforced when the E.U. Charter of Fundamental Rights enshrined data protection and privacy as fundamental rights,[302] which the ECJ used for a right-protective interpretation of the GDPR.[303] Jurisdictions without such constitutional safeguards and activist courts will operate under different conditions and will likely generate different outcomes.

[https://perma.cc/9CRF-4NPW]; Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, E.T.S. No. 108.

297. Directive 95/46/EC of Oct. 24, 1995, On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1995 O.J. (L 281/3) [hereinafter Data Protection Directive].

298. ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD ch. 5 (2020).

299. Paul Schwartz, *Global Data Privacy: The EU Way*, 94 N.Y.U. L. REV. 771, 779 (2019) (discussing diffusion of EU data protection law).

300. Lee A. Bygrave, *The "Strasbourg Effect" in Data Protection: Its Logic, Mechanics and Prospects in Light of the "Brussels Effect",* 40 COMPUT. L. & SEC. REV., Apr. 2021, at 1; Graham Greenleaf, *How Far Can Convention 108+ 'Globalise'? Prospects for Asian Accessions*, 40 COMPUT. L. & SEC. REV., Apr. 2021, at 2, § 1.2.

301. Graham Greenleaf & Bertil Cottier, *2020 Ends a Decade of 62 New Data Privacy Law*, 163 PRIV. L. & BUS. INT'L R. 24, 24 (2020).

302. Charter of Fundamental Rights of the European Union arts. 7–8, 2012 O.J. (C 326/391). The European Union is encouraging other jurisdictions to subscribe to this notion in its template for data governance provisions in trade agreements. *See* Draft Horizontal Provisions for Cross-Border Data Flows and for Personal Data Protection (in EU Trade and Investment Agreements), art. B.1, https://trade.ec.europa.eu/doclib/docs/2018/may/tradoc_156884.pdf [https://perma.cc/PF32-4WKA].

303. *See* Thomas Streinz, *The Evolution of European Data Law*, *in* THE EVOLUTION OF EU LAW 901, 910–12 (Paul Craig & Gráinne de Búrca eds., 3d ed. 2021).

Furthermore, as data protection and privacy laws proliferate—both in places where such legal frameworks had been previously absent and in places that had determined their existing laws to be in need of updating—legislators and the public will not only learn from each other, but also benefit from observing data controversies, debates, and litigation around the world and in different contexts. As a result, new features continue to appear.[304] The current extent of (and the potential for) future global harmonization of data protection and privacy laws ought not to be overstated. There are important differences between European-style data protection laws and the conceptions of data privacy that dominate the discourse in the United States and elsewhere.[305] We focus on data protection law in this Article, as this concept does seem to have more purchase globally,[306] but our main claims should apply *mutatis mutandis* to data privacy laws as well.

Data protection and privacy laws do not appear to be effective in challenging unequal control over data. This is partly by design and partly due to persistent underenforcement, even within the European Union, which is often perceived as the jurisdiction with the world's most stringent data protection law.[307] While data protection laws can achieve some rebalancing between individuals and data controllers by granting rights to the former against the latter, and may even achieve systemic change by requiring "data protection by design,"[308] data protection laws were not designed to address data inequality effectively. One design feature that further limits data protection law's

304. See, for example, a provision to permit disclosure of individuals' personal information without their knowledge or consent where disclosure is for "socially beneficial purposes." Bill C-11: An Act to Enact the Consumer Privacy Protection Act and the Personal Information and Data Protection Tribunal Act and to Make Related and Consequential Amendments to Other Acts, House of Commons, GOV'T OF CAN. (Dec. 2, 2020), https://www.justice.gc.ca/eng/csj-sjc/pl/charter-charte/c11.html [https://perma.cc/S8FL-6KB4].

305. *See generally* Paul M. Schwartz & Karl-Nikolaus Pfeifer, *Transatlantic Data Privacy Law*, 106 GEO. L.J. 115 (2017); Paul M. Schwartz & Daniel J. Solove, *Reconciling Personal Information in the United States and the European Union*, 102 CAL. L. REV. 877 (2014).

306. *See, e.g.*, Graham Greenleaf, *Global Data Privacy Laws 2021: Despite COVID Delays, 145 Laws Show GDPR Dominance*, 169 PRIV. L. & BUS. INT'L REP. 1, 5 (2021).

307. *See, e.g.*, Giovanni Buttarelli, Editorial, *The EU GDPR as a Clarion Call for a New Global Digital Gold Standard*, 6 INT'L DATA PRIV. L. 77, 77 (2016) (claiming that GDPR would raise the bar for data protection laws around the world). *But see* Margaret Taylor, *Data Protection: Threat to the GDPR's Status as 'Gold Standard'*, INT'L BAR ASS'N (Aug. 25, 2020), https://www.ibanet.org/article/A2AA6532-B5C0-4CCE-86F7-1EAA679ED532 [https://perma.cc/4DS6-XNRR] (raising concerns about GDPR's enforcement record).

308. *See* MIREILLE HILDEBRANDT, SMART TECHNOLOGIES AND THE END(S) OF LAW 220–21 (2016).

ability to confront data inequalities is its limited scope of application. This may seem counterintuitive for those who hail the comprehensiveness of the EU data protection regime. But even an umbrella framework such as the GDPR only applies to "personal data;"; "non-personal data" (e.g., highly aggregated data and anonymized data) is outside its scope of application.[309] Naturally, this conceptual demarcation creates an incentive to avoid the strictures of data protection law by shifting focus towards non-personal data. The binary distinction between personal and non-personal data is tenuous. Data is inherently relative: Data about myself might also be data about others.[310] Moreover, anonymized personal data may be vulnerable to reidentification.[311] Under conditions of vast ambient datafication, personally identifiable information may be derived from large and variegated datasets, even in instances when no "personal data" had been provided in the first place, which risks overstretching the GDPR's scope of application by turning the data protection law into a data "law of everything."[312]

The GDPR's focus on personal data is by design. Data protection law presumes that non-personal data does not raise questions of informational self-determination in the same way that personally identifiable information does. The fallacy of this assumption, however, can be illustrated by the prospect of synthetic data—artificially generated data with characteristics suitable for machine learning purposes, but without connection to any particular individual.

For some scholars, synthetic data and other "privacy preserving" technologies solve the "problem" of data protection law by allowing for datafication and data-driven decision-making without recourse to "personal data".[313] Note, however, that synthetic data

309. GDPR, *supra* note 184, art. 2(1); *see also* GDPR, *supra*, Recital 26 (non-applicability to anonymous data).

310. *See* Viljoen, *supra* note 287, at 603–07 (discussing a scenario where one individual's data necessarily implicates others as well).

311. *See, e.g.*, Michèle Finck & Frank Pallas, *They Who Must Not Be Identified – Distinguishing Personal from Non-Personal Data under the GDPR*, 10 INT'L DATA PRIV. L. 11, 15 (2020); Inge Graef, Raphael Gellert & Martin Husovec, *Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation*, 44 EUR. L. REV. 605, 610 (2019).

312. Nadezhda Purtova, *The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law*, 10 L., INNOVATION & TECH. 40, 41 (2018).

313. *See* Khaled El Emam, Lucy Mosquera & Richard Hoptroff, PRACTICAL SYNTHETIC DATA GENERATION: BALANCING PRIVACY AND THE BROAD AVAILABILITY OF DATA ch. 6 (2020). For a critical take from a privacy perspective, see Theresa Stadler, Bristena Oprisanu

aspires to reflect reality, thereby shaping and re-shaping the world as it is being represented, however imperfectly, through data. Concerns about uneven data-shaping power hence persist; the same is true for concerns about a possible concentration of synthetic data in the hands of few. Neither concern is being addressed by data protection law as it stands. What is true for synthetic data is true for other kinds of non-personal data being gathered by ever-expanding data infrastructures, especially through smartphones and other IoT devices.[314] Data about the environment, for example, while highly salient for policymaking to address climate change, is not governed by data protection law at all, unless it is tied to an individual. The individualistic approach to data protection law has thus rightly been recognized as a problem in confronting data-related harms that accrue collectively.[315]

A related design feature that curbs data protection law's ability to confront data inequality lies in the way in which an individual rights-based approach is being effectuated. Certain data protection rights enshrined in comprehensive data protection laws in the mold of the GDPR could, in theory, reduce data control asymmetries, at least with regard to personal data. For example, under the GDPR, data subjects have a relatively broad, though not unconditional, right to request the erasure of personal data.[316] If significant numbers of data subjects exercised this right, they would wrest control over personal data from data collectors. But they do not. This is a general weakness of an individual rights-based approach to data regulation. It depends on individuals' willingness and ability to exercise their rights. If they do not, the law remains ineffective. In the literature, the phenomenon that individuals profess a strong interest in data protection, but do not seem to act accordingly (as they routinely "give up" personal data with little regard to privacy), has been described as the "privacy paradox."[317] As Daniel Solove has demonstrated, the privacy paradox is actually not a paradox at all.[318] Managing one's privacy is a time-consuming and potentially nerve-wracking exercise. It is hence rational for individuals to proclaim an interest in data protection

& Carmela Troncoso, *Synthetic Data – Anonymisation Groundhog Day*, ARXIV (Dec. 11, 2020), https://arxiv.org/pdf/2011.07018.pdf [https://perma.cc/FG8S-UCLP].

314. *See* discussion *supra* Section I.C.

315. MARTIN TISNÉ, THE DATA DELUSION: PROTECTION INDIVIDUAL DATA ISN'T ENOUGH WHEN THE HARM IS COLLECTIVE 2 (2020), https://luminategroup.com/storage/1023/The-Data-Delusion—-July-2020.pdf [https://perma.cc/PPN9-65UL].

316. GDPR, *supra* note 184, art. 17.

317. Daniel J. Solove, *The Myth of the Privacy Paradox*, 89 GEO. WASH. L. REV. 1, 2 (2021).

318. *Id.* at 18–26.

generally, especially with regard to societal risks stemming from systemic surveillance, while not contributing to this effort individually by challenging such practices themselves.[319] This suggests the need for a shift towards a more collective and systemic enforcement of data protection law.

The GDPR has made some steps in this direction by improving the institutional infrastructure on which effective data protection law depends. EU law pioneered the idea of embedding data protection officers ("DPOs") within companies to affect the corporate culture towards data collection.[320] Independent data protection authorities ("DPAs") are tasked with investigating and sanctioning data protection violations.[321] By regulating algorithmic decision-making, the GDPR's novel regime provides an example for collaborative governance, in which companies' data protection impact assessments may offer systemic governance and suitable safeguards of individual rights implicated by algorithmic decision-making.[322] How effective these institutional upgrades will turn out to be remains to be seen. The enforcement record of the GDPR so far does not inspire confidence. In fact, underfunded DPAs struggled to fulfil their task to monitor resource-rich data controllers.[323] The continued existence of certain targeted advertising business models, long shown to be fundamentally incompatible with the GDPR, indicates how difficult it is to uproot

319. *Id.* at 24.

320. KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE 49–52 (2015). *But see* Ari Ezra Waldman, *Designing Without Privacy*, 55 HOUS. L. REV. 659, 681–89 (2018) (finding that privacy conceptions are narrow and limited and barely factor into the design of products).

321. *See* Streinz, *supra* note 303, at 913–14 (discussing institutional aspects of GDPR enforcement).

322. Margot Kaminski & Gianclaudio Malgieri, *Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations*, 11 INT'L DATA PRIV. L. 125, 127 (2020).

323. Even the otherwise largely self-congratulatory evaluation of the GDPR by the European Commission admits that:

> Given that the largest big tech multinationals are established in Ireland and Luxembourg, the data protection authorities of these countries act as lead authorities in many important cross-border cases and may need larger resources than their population would otherwise suggest. However, the situation is still uneven between Member States and not yet satisfactory overall.

Data Protection as a Pillar of Citizens' Empowerment and the EU's Approach to the Digital Transition – Two Years of Application of the General Data Protection Regulation, at 6, COM (2020) 264 final (June 24, 2020).

established data infrastructures deeply engrained in the digital economy.[324]

Even if individuals were willing and able to exercise their data protection rights, these rights might not go far enough to effectively challenge uneven control over personal data. The GDPR's novel right to data portability,[325] now being replicated in similar laws around the world, is illustrative of a rights-based approach that turned out to be too limited in scope and ignored important infrastructural dimensions. The right to data portability is an amalgam of data protection, competition, and telecommunication law rationales. Traditional data protection law recognized the right to access and erase one's personal data.[326] Data portability extends this idea to retrieve and/or transfer personal data from one data controller to the other in a way that can be understood as an expression of informational self-determination.[327] Competition law recognizes certain doctrines (e.g., the essential facilities doctrine) that grant mandatory access to certain categories of data when, without access to such data, market entry and effective competition become impossible.[328] Data portability can be understood in similar terms as confronting the pervasive network effects that have led to extreme platform concentration.[329] Competition and data protection law rationales for data portability may work in tandem when the reduction of switching costs leads to increased competition, on the assumption that competitors may seek to distinguish themselves through higher data protection standards. However, there is also a risk that the differing logics of competition and data protection law might interfere with one another.[330]

Data portability could in theory lead to a redistribution of personal data and potential de-concentration of infrastructural control over such data. Unfortunately, there is little evidence that this is actually happening. Two key reasons can be identified for data

---

324. *See, e.g.*, JOHNNY RYAN, SUBMISSION TO THE IRISH DATA PROTECTION COMMISSION: TWO YEARS ON FROM COMPLAINT TO THE IRISH DATA PROTECTION COMMISSION, THE RTB DATA BREACH IS THE LARGEST EVER RECORDED, AND APPEARS TO HAVE WORSENED 4 (2020).

325. GDPR, *supra* note 184, art. 20.

326. Data Protection Directive, *supra* note 297, art. 12(a)(b).

327. *See, e.g.*, Gabriela Zanfir, *The Right to Data Portability in the Context of the EU Data Protection Reform*, 2 INT'L DATA PRIV. L. 149, 152 (2012).

328. *See* discussion *infra* Section II.D.

329. The idea to prevent lock-in effects is also behind the idea, reflected in domestic and international telecommunications law, that one has a right to number portability to enable switching between different telecommunication services providers.

330. *See, e.g.*, Orla Lynskey, *Aligning Data Protection Rights with Competition Law Remedies? The GDPR Right to Data Portability*, 42 EUR. L. REV. 793, 795–96 (2017).

portability's limited impact. One concerns the scope of the right: the right to data portability under the GDPR is explicitly restricted to personal data that the data subject *provided* to the controller.[331] In other words, personal data that the data controller inferred about the data subject is not covered.[332] Moreover, as stated before, the value of personal data is highly contextual and relative. For this reason, the economic impact of data portability is often limited. For example, only the comment that a user provided on a social media platform has to be transferred to another platform—but not the picture provided by another user to which the comment was attached.[333] The second reason for data portability's limited impact (at least so far) is infrastructural. While the right to data portability under the GDPR requires that personal data be made available in a structured, commonly used, and machine-readable format—clearly recognizing that unstructured data in hard-to-access proprietary formats can constitute insurmountable obstacles to data access and use[334]—the GDPR fails to specify how the personal data actually ought to be transferred. This oversight is a missed opportunity to mandate and regulate private data transfer infrastructures (e.g., webAPIs) with public oversight, or to create alternative public data transfer infrastructures to ensure regulatability and generate interoperability.[335] In the absence of such infrastructural interventions, companies like Apple, Facebook, Google, Microsoft, and Twitter have developed an open source data transfer infrastructure—the Data Transfer Project—which they control, thereby indirectly affecting the effectiveness of the right to data portability.[336]

---

331. GDPR, *supra* note 184, art. 20(1).

332. The CCPA's right to access personal information in a portable and readily usable format can be construed as a data portability right, which—unlike GDPR—also covers inferred data. For a further comparison between GDPR and CCPA, see generally Anupam Chander, Margot E. Kaminski & William McGeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733 (2021).

333. *See* GABRIEL NICHOLAS & MICHAEL WEINBERG, DATA PORTABILITY AND PLATFORM COMPETITION: IS USER DATA EXPORTED FROM FACEBOOK ACTUALLY USEFUL TO COMPETITORS? 14 (2019), https://www.law.nyu.edu/sites/default/files/Data%20Portability%20and%20Platform%20Competition%20-%20Is%20User%20Data%20Exported%20From%20Facebook%20Actually%20Useful%20to%20Competitors.pdf [https://perma.cc/7ZEC-U8YX].

334. *See* discussion *supra* Section I.B.

335. *See* discussion *infra* Sections III.B and III.D.

336. *About Us*, DATA TRANSFER PROJECT, https://datatransferproject.dev [https://perma.cc/T5SR-XM8J]; *Google/Data-Transfer-Project*, GITHUB, https://github.com/google/data-transfer-project [https://perma.cc/Z5ZY-5HAY].

So far, we have discussed why data protection law has not been effective at challenging data inequality. In some instances, however, data protection law can even exacerbate data inequality.

One way that data protection law can come into conflict with attempts to redistribute control over personal data is when data controllers invoke data protection obligations to refuse data-sharing. The right to data portability, for example, is explicitly conditioned on not adversely affecting the data protection rights and freedoms of others,[337] a fact that is routinely stressed by platform companies in their discussion and practice of data portability.[338] Even though contemporary data protection laws recognize exemptions to facilitate data sharing for public benefit,[339] platform companies often adopt overly-restrictive interpretations of these exemptions and refuse to provide meaningful access to data for researchers studying the impact of platforms on peoples' lives and livelihoods.[340] Another way that data protection law may exacerbate data inequality concerns the increased compliance costs (assuming at least good-faith efforts at compliance) that may confer an inadvertent advantage to powerful and resource-rich accumulators of personal data. This question is conventionally only discussed under a competition and innovation policy framing, and views about a potential disparate impact of data protection law differ widely.[341] We suggest that it is a question worth asking not just from the perspective of competition and innovation policy, but also from a perspective of data inequality.[342]

---

337. GDPR, *supra* note 184, art. 20(4).

338. See, for example, the paper by Facebook's Chief Privacy Officer, ERIN EGAN, DATA PORTABILITY AND PRIVACY: CHARTING A WAY FORWARD 17 (Sept. 2019), https://about.fb.com/wp-content/uploads/2020/02/data-portability-privacy-white-paper.pdf [https://perma.cc/95DB-S2PF].

339. GDPR, *supra* note 184, arts. 85, 89.

340. Jef Ausloos, Paddy Leerssen & Pim ten Thije, *Operationalizing Research Access in Platform Governance: What to Learn From Other Industries?*, ALGORITHMWATCH (June 25, 2020), https://algorithmwatch.org/en/wp-content/uploads/2020/06/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf [https://perma.cc/4SMG-U8MX].

341. *See generally, e.g.*, Michal Gal & Oshrit Aviv, *The Competitive Effects of the GDPR*, 16 J. COMPETITION L. & ECON. 349 (2020) (arguing that GDPR limits competition in data markets, creating more concentrated market structures and entrenching the market power of those who are already strong; and that GDPR limits data sharing between different data collectors, thereby preventing the realization of data synergies which may lead to better data-based knowledge); Yafit Lev-Aretz & Katherine J. Strandburg, *Privacy Regulation and Innovation Policy*, 22 YALE J.L. & TECH. 256 (2020) (arguing that carefully designed privacy regulation can provide societally beneficial incentive structures for innovation).

342. See discussion *infra* Section II.D for discussion of what a competition law framing might overlook.

Having highlighted the inadequacy of data protection and privacy laws in resolving data inequality, we nonetheless remain sympathetic to its underlying objective of reclaiming the centrality of individual choice and autonomy against the over-intrusive power of corporate data collectors and processors. Still, we consider it worthwhile to contemplate whether the overemphasis on individual rights, even if beneficial in the short term, may in the long term legitimize the default position that datafication is both an acceptable and a desirable commercial activity, so long as certain concessions, in the form of enumerated rights, are made to individuals whose lives and environments are being datafied and affected by datafication. This concern echoes similar sentiments levelled by scholars like Samuel Moyn against human rights law (as ultimately not being effective at challenging rising economic inequality)[343] and Jessica Whyte, who has traced the co-constitution of human rights discourse (emphasizing individual freedoms against governmental intrusion) with the rise of the neoliberal project.[344] We flag these parallels to caution against a perception of extant data protection as an effective check on data inequality.[345] Data protection law operates under the assumption that if data controllers have legally acquired personal data, they may control that data as long as legitimate grounds for data processing exist. Certain limiting principles contained in data protection law—such as purpose limitation and data minimization—may have a dampening effect on data accumulation and repurposing and are at odds with the value proposition of "big data".[346] Safeguards against data collection (e.g., strict consent requirements for sensitive data) can amount to difficult-to-overcome obstacles to datafication. But ultimately, data protection law does not challenge concentrated control over data infrastructures, nor does it meaningfully constrain the power to datafy.[347] The COVID-19 pandemic revealed global corporations'

343. *See generally* SAMUEL MOYN, NOT ENOUGH: HUMAN RIGHTS IN AN UNEQUAL WORLD (2018). *But see generally* Gráinne de Búrca, 16 INT'L J. CONST. L. 1347 (2018) (book review).

344. *See* JESSICA WHYTE, THE MORALS OF THE MARKET: HUMAN RIGHTS AND THE RISE OF NEOLIBERALISM ch. 3 (2020).

345. *See* Angela Daly, *Neo-Liberal Business-As-Usual or Post-Surveillance Capitalism With European Characteristics? The EU's General Data Protection Regulation in a Multi-Polar Internet*, *in* COMMUNICATION INNOVATION AND INFRASTRUCTURE: A CRITIQUE OF THE NEW IN A MULTIPOLAR WORLD (Rolien Hoyng & Gladys Pak Lei Chong eds., forthcoming 2021), https://ssrn.com/abstract=3655773 [https://perma.cc/GY6A-FEWP].

346. *See, e.g.*, Tal Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1004–12 (2017).

347. There is also a risk that, through practices of implementation, data protection law is being transformed into a mere compliance exercise ("check list") or a problem to be solved

control over data infrastructures when Apple and Google collaborated to make available within their mobile operating systems Bluetooth-powered COVID-19 contact tracing apps sanctioned by public health authorities, but refused to adapt their systems to enable the centralized collection of data favored by French and British health authorities. While largely supported by privacy advocates, Apple's and Google's refusal laid bare that data protection law does not confront centralized corporate control over large scale data-generating infrastructures—in this case, the operating systems running on billions of mobile devices.[348]

In the next Section, we discuss the potential and limits of competition law as a supplement or corollary to data protection law to address "platform power."

D. Regulating Platform Power

As we have seen, certain platform companies have created expansive data infrastructures.[349] The rise of platform power has led to increased regulatory and scholarly scrutiny. This discourse has largely focused on: (1) laws protecting platforms from liability for hosted content (within the relatively novel field of intermediary liability law);[350] (2) regulations specifically aimed at regulating relationships between platforms, businesses, and consumers (either through dedicated platform regulation or consumer protection law); and (3) tools of antitrust and competition law. While the discussion about platform liability laws is often framed through the lens of communicative freedoms and related harms ranging from copyright violations to hate speech, antitrust and competition laws have focused on platform companies as dominant market actors and evaluated their impact on dependent commercial actors, potential competitors, and consumers. Platform regulation consists of a complex and disparate set of laws deeply intertwined with the rise of platform companies and

---

rather than a philosophy to be embraced. This, in turn, may contribute to the notion that mere compliance with data protection law is sufficient from a public policy perspective. For the opposite approach, see generally Mireille Hildebrandt, *Primitives of Legal Protection in the Era of Data-Driven Platforms, 2* GEO. L. & TECH. REV. 252 (2018); HILDEBRANDT, *supra* note 308.

348. Michael Veale, *Sovereignty, Privacy, and Contact Tracing Protocols*, *in* DATA JUSTICE AND COVID-19: GLOBAL PERSPECTIVES 34, 37–39 (Linnet Taylor et al. eds., 2020).

349. *See* discussion *supra* Section I.C.

350. The OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY (Giancarlo Frosio ed., 2020) is testament to the emergence of such a field.

informational capitalism.[351] For purposes of this Article, we focus on the ways in which platform regulation has enabled data inequality. We also seek to find avenues through which such regulation could mitigate or reduce data inequality going forward. Even though platform regulation tends to be more attuned to infrastructural dimensions than other regulatory approaches, we expose significant limitations inherent in extant approaches to platform regulation.

We first turn to established intermediary liability laws and newly emerging platform regulation.[352] We then address the evolving debate in antitrust and competition law around growing platform power, highlighting their respective salience for questions of data inequality.[353] Each of these legal frames asks its own unique sets of questions that all touch on data inequality but rarely focus on it. In contrast, we seek to foreground the salience of platform regulation for data inequality and to expose the limitations of extant approaches.

Liability shields for user-generated content have long been portrayed as critical for internet freedom.[354] The United States pioneered this regulatory approach for user-generated content. Under the liability shield introduced by the Communications Decency Act of 1996, providers of platforms are not to be treated as publishers or speakers of information provided by others.[355] The Digital Millennium Copyright Act of 1999 created a notice and takedown regime which protects platforms against liability for copyright protected content.[356] These legal mechanisms have been replicated, albeit in more limited form, in jurisdictions around the world.[357] They also feature in recent U.S.-designed "digital trade" agreements.[358] However, criticism is mounting that liability shields for user-generated content and their expansive interpretation by U.S. courts accord a subsidy for platform power by effectively disabling liability as a legal

351. *See* COHEN, *supra* note 14, at 108–37.

352. *See* text accompanying notes 354—362.

353. *See* text accompanying notes 363—406.

354. *See generally CDA 230: The Most Important Law Protecting Internet Speech*, ELEC. FRONTIERS FOUND., https://www.eff.org/issues/cda230 [https://perma.cc/ZQ7L-538R]; JEFF KOSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET (2019).

355. Communications Act of 1934, 47 U.S.C. § 230 (1934) (as amended).

356. Digital Millennium Copyright Act of 1998, 17 U.S.C. § 512 (1998).

357. *See, e.g.*, Directive 2000/31/EC, of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market, 2000 O.J. (L 178/1), arts. 12–15. A global overview over intermediary liability laws is available at STANFORD WORLD INTERMEDIARY LIABILITY MAP, https://wilmap.law.stanford.edu/ [https://perma.cc/5THN-4FDN].

358. *See, e.g.*, USMCA, *supra* note 276, art. 19.17.

technology inducing appropriate corporate behavior.[359] The main focus of this debate, at least in the United States, is on content-related harms and the platforms' responsibility for enabling such harms.[360] One tangent within the debate concerns the question of whether small or big platforms benefit more, and in what ways, from intermediary liability protection.[361] From the perspective of data inequality that this Article takes as its focus, intermediary liability seems only indirectly related to uneven control over data and unequal power to datafy.[362]

However, platform companies have also emerged as targets for dedicated platform regulation that transcends the debate around intermediary liability. One such example is the European Union's regulation on platform-to-business (p2b) relations.[363] This type of platform regulation singles out a particular type of intermediary: namely, platforms that allow business users to offer goods or services to consumers.[364] The regulation supplements the contractual relationships between platforms and business users. One of its main interventions is to demand a certain degree of transparency to guard against information asymmetries between platforms and business users.[365] Nevertheless, the regulation shies away from demanding the

359. *See, e.g.*, COHEN, *supra* note 14, at 97–101.

360. *See, e.g.*, Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform* (B.U. Sch. of L., Pub. L. & Legal Theory Rsch. Paper No. 20-8 2000), https://ssrn.com/abstract=3532691 [https://perma.cc/R3SG-BADF]; Olivier Sylvain, *Discriminatory Designs on User Data*, KNIGHT FIRST AMEND. INST. EMERGING THREATS ESSAY SERIES (Apr. 1, 2018), https://knightcolumbia.org/content/discriminatory-designs-user-data [https://perma.cc/3GVP-22GE].

361. *See, e.g.*, Eric Goldman, *Want to Kill Facebook and Google? Preserving Section 230 Is Your Best Hope*, BALKINIZATION (June 3, 2019), https://balkin.blogspot.com/2019/06/want-to-kill-facebook-and-google.html [https://perma.cc/L2WB-SCUJ].

362. How platform companies would have developed in the absence of protections against intermediary liability over the last quarter century is a difficult to answer hypothetically. Conversely, it is difficult to predict what kind of reforms to Section 230 might effectively curb the dominance of large platform companies. Still, the existence of protections afforded by Section 230 has likely empowered platform companies to deflect challenges to their business models and impacted the distribution of power to datafy. *See also* discussion *supra* Section I.D.

363. Regulation 2019/1150, of the European Parliament and of the Council of 20 June 2019 on Promoting Fairness and Transparency for Business Users of Online Intermediation Services, 2019 O.J. (L 186/57) [hereinafter p2b Regulation].

364. *See id.*, art. 2(2) (for the composite definition of "online intermediation services" in the p2b regulation).

365. For further ideas and even more assertive ways of demanding transparency from data infrastructure controllers, see discussion *infra* Section III.C.

sharing of the data that e-commerce platforms generate about business and consumer behavior. Instead, it requires platforms to disclose in their terms and conditions which business or consumer data is being generated and who has (or does not have) access to it.[366] The idea seems to be that such disclosure might enable businesses to negotiate more favorable terms for access to data from platforms. Although moves to increase transparency should be viewed favorably and, as we argue in Part III, should be pursued more aggressively, it seems unlikely that such an intervention in itself is sufficient to effectively redistribute control over data through contractual means, given the power asymmetries between platforms and businesses and the increasing business dependency on e-commerce platforms, which has been exacerbated by the COVID-19 pandemic.[367]

Dedicated platform regulation to address information asymmetries is a relatively recent phenomenon. The European Union's ambitious digital single market strategy includes proposals not only to reform the liability regime established under the e-commerce directive but also to provide a new regulatory framework for digital services through the proposed Digital Services Act (DSA) and the Digital Markets Act (DMA).[368] Both instruments are meant to complement each other, but they adopt different concepts and logics to achieve their respective regulatory objectives. The DSA is designed around different categories of "intermediary services" and regulates providers' respective liability and due diligence obligations, as well as their enforcement.[369] The DMA singles out certain "core platform services" as gatekeepers on which additional obligations are being imposed, including prohibitions to engage in certain business

366. *See* p2b regulation, *supra* note 363, art. 9 (misleadingly labelled "access to data").

367. OECD, E-COMMERCE IN THE TIMES OF COVID-19, at 6–7 (2020), https://read.oecd-ilibrary.org/view/?ref=137_137212-t0fjgnerdb&title=E-commerce-in-the-time-of-COVID-19 [https://perma.cc/6JRF-HXJ7]. For ideas how to counter asymmetric control over data through pooling strategies, see discussion *infra* Section III.D.

368. The European Commission refers to these together referred as the Digital Services Package. *The Digital Services Act Package*, EUROPEAN COMM'N, https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package [https://perma.cc/SM8H-U48A].

369. *Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC*, COM (2020) 825 final (Dec. 15, 2020) [hereinafter DSA Proposal]. On April 23, 2022, the European Parliament and Council reached a political agreement on the DSA. After the text has been finalized and formally approved, the final version of the DSA will be published in the EU's Official Journal.

practices.[370] As under the p2b regulation, transparency obligations feature prominently. The DSA focuses on transparency regarding illegal content and platforms' content moderation practices.[371] The DMA seeks to make more visible for advertisers and publishers how gatekeeper platforms set their prices for advertising services.[372]

Both laws go beyond transparency obligations. The DMA subjects "gatekeepers"[373] to various requirements designed to facilitate or mandate data sharing with businesses and end users.[374] Institutions tasked with enforcing the DSA can request access to any data necessary to monitor and assess compliance with the DSA from "very large" online platforms.[375] The European Commission may also request access to databases when conducting market investigations or enforcing the DMA.[376] The DMA proposal thus foresees a shift towards more centralized, robust, and intrusive enforcement compared

370. *Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*, COM (2020) 842 final (Dec. 15, 2020) [hereinafter DMA Proposal]. "Core platform services" are defined as any digital service included in the exhaustive list enumerated by Article 2.2 of the DMA: online intermediation services, online search engines, online social networking services, video-sharing platform services, number-independent interpersonal communication services, operating services, cloud computing services, and advertising services. On March 24, 2022, the European Parliament and Council reached a political agreement on the DMA, and an unofficial text of the DMA was leaked on April 13, 2022. After the text of the DMA has been finalized and formally approved, the final version of the DMA will be published in the EU's Official Journal.

371. *See, e.g.*, DSA Proposal, *supra* note 369, art. 13.

372. *See, e.g.*, DMA Proposal, *supra* note 370, art. 5(g), recitals 42, 53. Note also the European Commission's guidelines on ranking transparency pursuant to the p2b Regulation. European Comm'n, Commission Notice, Guidelines on Ranking Transparency Pursuant to Regulation (EU) 2019/1150 of the European Parliament and of the Council, 2020 O.J. (C 424/1).

373. Gatekeepers are being defined as "provider[s] of core platform services" that have a significant impact on the internal market," "serve[] as important gateway[s] for business users to reach end users," and "enjoy[] an entrenched and foreseeably durable position." *See* DMA Proposal, *supra* note 370, art. 3.

374. *See, e.g.*, *id*. art. 6.1(g) (giving advertisers and publishers a right to request, free of charge, access to performance measuring tools and ad inventory verification information); *id*. art. 6(h) (reinforcing the GDPR's right to data portability, discussed *infra* Section II.C., by requiring "tools for end users to facilitate the exercise of data portability, in line with [GDPR], including by the provision of continuous and real-time access.").

375. DSA Proposal, *supra* note 369, art. 31; *id.* art. 25 (defining "[v]ery large online platforms" as those providing services to forty-five million (or more) average monthly active users within the EU).

376. DMA Proposal, *supra* note 370, art. 19.

to the GDPR.[377] In what form (if at all) these proposals will be adopted remains to be seen. What is relevant for purposes of this Article is that the DSA and DMA proposals can be seen as further evidence towards dedicated yet differentiated platform regulation that complement and transcend established intermediary liability regimes.

In recognizing the central (infrastructural) role of "very large" online platforms (under the DSA) and acknowledging the widespread tracking and profiling activities of many large platforms (under the DMA), the European Commission focuses its regulatory attention directly on the role of platform companies in Europe's digital economy. At the same time, the Commission's proposals neither take issue with the amassing of data[378] as such, nor do they question the desirability of ever-increasing datafication.[379] The increased regulatory burdens under a potential DSA or DMA could have a dampening effect on certain types of data generation wherever not generating data is a viable option.[380]

The proposed DMA explicitly prohibits "gatekeepers" from combining personal data sourced from core platform services with other services offered by the gatekeeper, but the prohibition can be overcome if end users are being provided with a specific choice and provide consent under the GDPR.[381] In similar fashion, in February 2019, German antitrust authorities required Facebook Inc. (now Meta Platforms Inc.) to cease aggregations of personal data across its platforms (Facebook, Instagram, and WhatsApp) in the absence of GDPR-compliant consent.[382] The *Bundeskartellamt* justified this

377. *See, e.g.*, JOAN BARATA ET AL., UNRAVELLING THE DIGITAL SERVICES ACT PACKAGE, EUR. AUDIOVISUAL OBSERVATORY 97 (2021); *see also* Pierre Larouche & Alexandre de Streel, *The European Digital Markets Act: A Revolution Grounded on Traditions*, J. EUR. COMP. L. & PRAC. 542 (2021).

378. The DMA states that comprehensive tracking and profiling of end users online as such is not necessarily an issue if done in a controlled and transparent manner, in respect of privacy, data protection and consumer protection. DMA Proposal, *supra* note 370, at 1 n.1.

379. Indeed, the European Commission's European strategy for data celebrates growing data volumes around the world. On the questionable relevance of mere quantitative metrics about data, see Strasser & Edwards, *supra* note 27.

380. *Cf.* Gal & Aviv, *supra* note 341 (analyzing the impact of GDPR on companies' incentives to collect, process, or share data). A comparable analysis for the DSA and DMA is beyond the scope of this Article.

381. DMA Proposal, *supra* note 370, art. 5(a).

382. Bundeskartellamt [Federal Cartel Office] Feb. 6, 2019, B6-22/16, (Ger.) https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Entscheidungen/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=5 [https://perma.cc/TZR5-KFB9]; Bundesgerichtshof [BGH] [Federal Court of Justice] June 23, 2020, KVR 69/19 (upholding the law in preliminary proceedings by the German Federal Court of Justice).

intervention within the established categories of European competition law, alleging that Facebook had abused its dominant market position in a way that vitiated users' pro-forma consent. The proposed DMA, in contrast, would impose a non-aggregation obligation outright against all "gatekeepers."[383] The regulatory prohibitions under the DMA proposal are thus illustrative of the differences between regulatory intervention and competition law enforcement.

Competition authorities in the European Union have been conducting extensive investigations into the conduct of platform companies, triggering a debate about the proper scope and purpose of competition law.[384] Antitrust authorities elsewhere, including in the United States, have ramped up their investigations into platform companies' conduct.[385] Questions of access to and control over data have featured prominently in these investigations and debates. Indeed, antitrust and competition law seem conceptually better equipped to address systemic issues of data inequality than property rights or otherwise individual rights-based approaches. Yet, as we shall see, there are also important limitations inherent in antitrust and competition law that need to be acknowledged. As with data protection and privacy law, there are commonalities, but also

383. DMA Proposal, *supra* note 370, art. 5(a).

384. *See, e.g.*, Autorité de la Concurrence & Bundeskartellamt, Competition Law and Data (May 10, 2016), https://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2 [https://perma.cc/YF9U-6BKT]. For the debate on competition law, see generally STIGLER COMMITTEE ON DIGITAL PLATFORMS, FINAL REPORT (2019); UNLOCKING DIGITAL COMPETITION: REPORT OF THE DIGITAL COMPETITION EXPERT PANEL (Mar. 2019); JACQUES CRÉMER, YVES-ALEXANDRE DE MONTJOYE & HEIKE SCHWEITZER, COMPETITION POLICY FOR THE DIGITAL ERA (2019).

385. MAJORITY STAFF REPORT, *supra* note 106. In October 2020, the U.S. Department of Justice, together with several states, filed a lawsuit against Google alleging that the company unlawfully maintained monopolies in the markets for general search services, search advertising, and general search text advertising in the United States through anticompetitive and exclusionary practices. Press Release, U.S. Dep't of Justice, Justice Department Sues Monopolist Google for Violating Antitrust Laws (Oct. 20, 2020), https://www.justice.gov/opa/pr/justice-department-sues-monopolist-google-violating-antitrust-laws [https://perma.cc/TH92-U97V]. A trial is set for September 12, 2023. *Id.* In December 2020, the Federal Trade Commission sued Facebook, alleging that the company illegally maintained a monopoly for personal social networking through anticompetitive conduct. Press Release, Fed. Trade Comm'n, FTC Sues Facebook for Illegal Monopolization, https://www.ftc.gov/news-events/press-releases/2020/12/ftc-sues-facebook-illegal-monopolization [https://perma.cc/CX76-XQLK]. On August 19, 2021, the FTC filed an amended complaint adding additional details in support of its claims. First Amended Complaint for Injunctive and Other Equitable Relief, Federal Trade Comm'n v. Facebook, Inc., No. 1:20-cv-03590-JEB (D.D.C. Aug. 19, 2021) (No. 75-1).

important differences, between the respective legal regimes in the United States and Europe, with most jurisdictions elsewhere gravitating towards the E.U. approach to competition law.[386]

The platform power that U.S.-based corporations have accumulated has emerged as one main target for increased antitrust scrutiny in the United States.[387] Such inquiries grapple with the questions of concentrated control over data with which our Article is engaged. However, neither data accumulation as such, nor the power to datafy, is their main concern. Even under the resurgent yet highly contested "Neo Brandeisian" framing of antitrust law as advocated by Lina Khan and others, there needs to be some kind of anti-competitive conduct to justify interventions by antitrust authorities.[388] Even in this reimagined form, the harm with which antitrust law remains concerned is the potential harm to competition caused by concentrated market power.[389] This concern may overlap to a significant extent with the data inequality concerns outlined above, especially when market power enables platform companies to attain control over expansive

386. Anu Bradford et al., *The Global Dominance of European Competition Law Over American Antitrust Law*, 16 J. EMPIRICAL LEGAL STUD. 731, 734 (2019) (finding that the EU's competition laws have been more widely emulated than the US's competition laws); *see also* Thomas K. Cheng, *Convergence and Its Discontents: A Reconsideration of the Merits of Convergence of Global Competition Law* 12 CHI. J. INT'L L. 433 (2012) (critiquing the global convergence of competition laws).

387. *See supra* note 385 and accompanying text; *see also* Lina Khan, *Sources of Tech Platform Power*, 2 GEO. L. TECH. REV. 325, 331–32 (2018).

388. What is known as antitrust law in the United States harkens back to the Sherman Antitrust Act of 1890, passed in response to the concentration of corporate power during the Gilded Age in the United States. *See* TIM WU, THE CURSE OF BIGNESS: ANTITRUST IN THE NEW GILDED AGE 10–11 (2018). From the 1970s onwards, the so-called "Chicago School" reoriented U.S. antitrust law towards a law and economics driven analysis of "consumer welfare." *See generally* Herbert Hovenkamp & Fiona Scott Morton, *Framing the Chicago School of Antitrust Analysis*, 168 UNIV. PA. L. REV. 1843 (2020). More recently, this shift has been criticized for casting aside certain political dimensions of antitrust law. *See, e.g.*, Ariel Katz, *The Chicago School and the Forgotten Political Dimension of Antitrust Law*, 87 UNIV. CHI. L. REV. 413, 416–17 (2020). For a more general critique of the economic efficiency paradigm, see Jedediah Britton-Purdy et. al, *Building a Law-and-Political-Economy Framework: Beyond the Twentieth-Century Synthesis*, 129 YALE L.J. 1784, 1800–02 (2020). On the "Neo Brandeisians," see generally Lina Khan, *The New Brandeis Movement: America's Antimonopoly Debate*, 9 J. EUR. COMPETITION L. & PRAC. 131 (2018).

389. *See, e.g.*, Lina Khan, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 710 (2017) (arguing that the predominant framework in U.S. antitrust—specifically its "pegging competition to 'consumer welfare,' defined as short-term price effects—is unequipped to capture the architecture of market power in the modern economy" and risks missing potential harms to competition).

data infrastructures.[390] But certain data inequality concerns will remain out of the focus of antitrust analysis. For starters, data concentration is not always a function of market concentration and may in fact occur across markets, even transnationally. Moreover, data concentration as such is not a concern. Data concentration is only seen as an enabling factor that might lead to a monopolistic position, which is in itself not a problem from an antitrust law perspective unless a monopolist uses its position to engage in anticompetitive conduct. Likewise, when scrutinizing transactions through which companies' data or data infrastructures are acquired, antitrust law only intervenes if there are anticompetitive effects.[391] In other words, antitrust law is only concerned with distributional effects when they are the result of anticompetitive conduct. Crucially, however, harms of data inequality are not necessarily caused by or identical to harms to competition. Moreover, the non-existence or non-availability of data—for example, data that is necessary to measure sustainable development or to craft well-tailored public policies—is not an issue that antitrust law is designed to address at all.[392]

Competitors, however, could have a chance to gain access to data and data infrastructures under U.S. antitrust law, if the United States were to revive, renew, and expand the "essential facilities doctrine"[393]—despite U.S. Supreme Court case law that has gradually curtailed its potential scope of application and that seems, at least so far, largely oblivious to digital platform dynamics and related

390. *See* discussion *supra* Section I.C.

391. *See, e.g.*, Competitive Impact Statement at 17–18, United States v. Google Inc., 2011 WL 2444825 (D.D.C. Oct. 5, 2011) (No. 1:11-cv-00688). The court approved Google's acquisition of ITA Software, thereby allowing Google to acquire data and algorithms used to combine and parse flight information from airlines, including pricing and availability data. *See* United States v. Google, Inc., 2011 U.S. Dist. LEXIS 124151, at *43–44 (D.D.C. Oct. 5, 2011). The court imposed a time limit remedy which required Google to license ITA's data infrastructure to other websites for a period of five years. *Id.* at *22.

392. *See* Orla Lynskey, *Regulating 'Platform Power'* 4–9 (London Sch. of Econ. and Pol. Sci., L., Soc'y and Econ. Working Paper No. 1/2017, 2017) (arguing that platform power is over- and under-inclusive and that certain issues need to be addressed outside a competition law framework); *see also* Ariel Ezrachi, *EU Competition Law Goals and the Digital Economy* 1 (Oxford Legal Stud., Rsch. Paper No. 17/2018, 2018) (asking: is this a competition problem?).

393. For arguments in favor of such a move, see generally Nikolas Guggenberger, *Essential Platforms*, 24 STAN. TECH. L. REV. 237 (2021). *See generally* Zachary Abrahamson, *Essential Data*, 124 YALE L.J. 867 (2014); Brett M. Frischmann & Spencer Weber Waller, *Revitalizing Essential Facilities*, 75 ANTITRUST L.J. 1 (2008). The basic idea behind the doctrine is that a monopolist who owns a facility which is essential to its competitors must grant reasonable use of that facility under certain conditions. Frischmann & Waller, *supra* at 4.

harms.[394] Under this doctrine, a monopolist who owns "a facility essential to other competitors" must grant reasonable use of that facility under certain conditions.[395] The European Union's analogue to the "essential facilities doctrine" seems better equipped to address this particular type of data inequality (i.e. lack of access to data and data infrastructures by competitors) because the relevant thresholds to establish abuse of a dominant market position seem more favorable to competitors seeking access than the corresponding thresholds under U.S. antitrust law.[396] In the absence of competition law claims, the European Union has established certain access to data rights beyond personal data portability for businesses and consumers in certain sectors.[397] Note, however, that such mandatory data sharing may perpetuate the initial determinations made under recourse to the data infrastructure controller's power to datafy unless the regulation itself mandates what kind of data has to be generated, retained, and shared.[398] Crafting such regulatory intervention is complicated, and the European Union has so far refrained from far-reaching mandatory data sharing, settling instead for sectoral access to data regimes (e.g., for electricity data in the context of "smart meters").[399] Targeted

394. Verizon Commc'ns Inc. v. Curtis V. Trinko, LLP, 540 U.S. 398, 410–12 (2004) (holding that antitrust remedies are not available for regulated industries); Ohio v. American Express Co., 138 S. Ct. 2274, 2274 (2018) (upholding the restrictive measures of a credit card company for lack of harm); *see* Tim Wu, *The* American Express *Opinion, the Rule of Reason, and Tech Platforms*, 7 J. ANTITRUST ENF'T 104, 117 (2019).

395. Abbott B. Lipsky, Jr. & J. Gregory Sidak, *Essential Facilities*, 51 STAN. L. REV. 1187, 1190–91 (1999).

396. For an in-depth analysis of EU competition law from an essential facilities perspective, see generally INGE GRAEF, EU COMPETITION LAW, DATA PROTECTION AND ONLINE PLATFORMS: DATA AS ESSENTIAL FACILITY (2016); Inge Graef, *When Data Evolves into Market Power–Data Concentration and Data Abuse Under Competition Law*, *in* DIGITAL DOMINANCE: THE POWER OF GOOGLE, AMAZON, FACEBOOK, AND APPLE (Martin Moore & Damian Tambini eds., 2018).

397. *See, e.g.*, Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on Common Rules for the Internal Market for Electricity and Amending Directive 2012/27/EU, 2019 O.J. (L 158/125) (establishing guidelines on transparency in consumer access to energy consumption data).

398. Transparency obligations regarding data infrastructures, including disclosure of choices made and methodologies involved in data generation is another regulatory avenue. *See* discussion *infra* Section III.C.

399. For a summary of EU law on sharing of non-personal data, see Support Centre for Data Sharing, *Analytical Report on EU Law Applicable to Sharing of Non-Personal Data*, DG CONNECT (Jan. 24, 2020), https://eudatasharing.eu/index.php/legal-aspects/report-eu-law-applicable-sharing-non-personal-data [https://perma.cc/G8KZ-MTCM].

access to data rights is also common in other jurisdictions (e.g., for access to automotive data for repair shops).[400]

The corollary to provisions that enable access to and transfer of data (i.e. data portability) are interventions that require interoperability between data infrastructures.[401] Interoperability can be facilitated through a private standard-setting organization (e.g., with regard to data standards),[402] as required under antitrust and competition law[403] or mandated by regulators.[404] The proposed DMA takes the latter route by requiring "gatekeepers" to allow business users and providers of ancillary services access to and interoperability with the operating system, hardware, or software features used by the gatekeeper for its own ancillary services.[405] Such a regulatory intervention would diminish the infrastructural control of developers of operating systems and device manufacturers, such as Apple. While envisaged as complementary to competition law remedies, including the essential facilities doctrine, the proposed DMA shares with competition law a framing that is chiefly concerned with gatekeepers' impact on innovation and competition.[406] Although the proposed

400. The Parliament of the Commonwealth of Australia, Competition and Consumer Amendment (Motor Vehicle Service and Repair Information Sharing Scheme) Bill 2020, Exposure Draft, https://treasury.gov.au/sites/default/files/2020-12/c2020-128289-exposure-draft.pdf [https://perma.cc/K5UP-THP9]. The draft bill was open for public comments until January 31, 2021.

401. *See, e.g.*, CRÉMER, MONTJOYE & SCHWEITZER, *supra* note 384, at 58–59 (distinguishing between protocol interoperability, data interoperability, and full protocol interoperability); *see also* Przemysław Pałka, *The World of Fifty (Interoperable) Facebooks*, 51 SETON HALL REV. 1193, 1228–35 (2021).

402. *See, e.g.*, Michal S. Gal & Daniel L. Rubinfeld, *Data Standardization*, 94 N.Y.U. L. REV. 737, 749–54 (2019).

403. *See, e.g.*, Chris Riley, *Unpacking Interoperability in Competition*, 5 J. CYBER POL'Y 94, 94–96 (2020).

404. *See, e.g.*, IAN BROWN, OPENFORUM ACADEMY, INTEROPERABILITY AS A TOOL FOR COMPETITION REGULATION 47 (Nov. 2020), https://osf.io/preprints/lawarxiv/fbvxd/ [https://perma.cc/B4VT-QLTU].

405. DMA Proposal, *supra* note 370, art. 6(f), recital 52.

406. *See, e.g.*, *id*., recital 54:

> Gatekeepers benefit from access to vast amounts of data that they collect while providing the core platform services as well as other digital services. To ensure that gatekeepers do not undermine the contestability of core platform services as well as the innovation potential of the dynamic digital sector by restricting the ability of business users to effectively port their data, business users and end users should be granted effective and immediate access to the data they provided or generated in the context of their use of the relevant core platform services of the gatekeeper, in a structured, commonly used and machine-readable format. This should apply also to any other data at different levels of aggregation that may be necessary to effectively enable such portability. It should also be ensured

DMA is not challenging data inequality as such, it may still have positive effects in this regard by making visible and redistributing infrastructural control.

While the focus on competition and innovation is a commonality, regulatory ex ante intervention, as proposed under the DMA, differs from ex post remedies under competition law insofar as the latter requires engaging with core competition law concepts, such as market definition and market dominance. The ways in which antitrust and competition regulators in the United States and the European Union have analyzed mergers in the tech sector (at least so far) are illustrative of the resulting blind spots. The acquisition of additional data infrastructures as such is of no concern from an antitrust or competition law perspective unless it leads to a monopoly or dominant market position with adverse effects on competition. Such prognosis is, of course, dependent on accurate information about what the companies plan to do with the acquired infrastructures (or the data). This was not the case when Facebook misled the European Commission about the possibility of aggregating Facebook data with WhatsApp data.[407] Even without outright deception, the estimation of prognostic use is susceptible to miscalculations. In some merger cases, companies have made data sharing concessions to ease data concentration concerns.[408] In many other cases, however, companies managed to survive merger control scrutiny because the authorities were only focused on the effects of the merger in particular markets and were not interested in broader concerns around increased data concentration, including those that might have resulted from service integration across different markets.[409] This purist view of competition law can lead to a systemic overlooking of large-scale data

---

> that business users and end users can port that data in real time effectively, such as for example through high quality application programming interfaces. Facilitating switching or multi-homing should lead, in turn, to an increased choice for business users and end users and an incentive for gatekeepers and business users to innovate.

407. European Commission Press Release IP/17/1369, Mergers: Commission Fines Facebook €110 million for Providing Misleading Information about WhatsApp Takeover (May 18, 2017).

408. *See, e.g.*, Commission Decision, Case COMP/M.4854 TomTom/TeleAtlas, 2008 O.J. (C237/53); Commission Decision, Case COMP/M.6314 Telefonica/Vodafone/EE, 2012 O.J. (C66/122); Commission Decision, Case COMP/M.7023 Publicis/Omnicon, 2014 O.J. (C84/112).

409. *See, e.g.*, Commission Decision, Case COMP/M.7217 Facebook/WhatsApp, 2014 (7239 final) (focused exclusively on a potential strengthening of Facebook's position in online advertising and dismissed privacy-related concerns as outside the scope of EU competition law).

accumulation across complex data infrastructures.[410] For some, this limited and clearly-delineated scope of competition law analysis is preferable to an expansion in contradiction to established tenets of the regime.[411] For others, a recalibration of the established regimes is unavoidable due to the versatility and variability of data thus advocating for integration of data protection and privacy law.[412] These different views have implications for competition law's suitability to address data inequality.

One final limitation of antitrust and competition law that is not always fully acknowledged deserves to be foregrounded: Despite some structures for international coordination and a global agenda to install antitrust and competition law regimes in jurisdictions around the world,[413] antitrust and competition law remain fundamentally concerned with the anticompetitive effects on domestic markets. That is, U.S. antitrust law is concerned with the U.S. market, European competition law focuses on the European market and so on. This statement is true regardless of the occasional invocation of "extraterritorial" jurisdiction when anticompetitive conduct occurs outside their jurisdiction but materializes within their jurisdiction (under the effects doctrine), as the analysis remains confined to domestic effects and ignores global implications.[414] The U.S. Supreme Court has explicitly dismissed the idea that U.S. antitrust law should remedy anticompetitive conduct abroad.[415] Even if antitrust and

410. *See, e.g.*, Reuben Binns & Elettra Bietti, *Dissolving Privacy, One Merger at a Time: Competition, Data, and Third Party Tracking*, 36 COMPUT. L. & SEC. REV. 2020 (analyzing data accumulation qua user tracking).

411. *See generally* NICOLAS PETIT, BIG TECH AND THE DIGITAL ECONOMY: THE MOLIGOPOLY SCENARIO (2020) (arguing for regulation, not competition, as the appropriate tool to address non-competition harms).

412. *See* Orla Lynskey, *Grappling with "Data Power": Normative Nudges from Data Protection and Privacy*, 20 THEORETICAL INQUIRIES IN L. 189 (2019).

413. *See, e.g.*, Eleanor M. Fox & Amedeo Arena, *The International Institutions of Competition Law: The Systems' Norms*, *in* THE DESIGN OF COMPETITION LAW INSTITUTIONS (Eleanor Fox & Michael Trebilcock eds., 2012) (discussing the WTO and other international institutions).

414. *See, e.g.*, Eleanor M. Fox, *National Law, Global Markets, and* Hartford*: Eyes Wide Shut* 68 ANTITRUST L.J. 73 (2000) (criticizing the U.S. Supreme Court decision in Hartford Fire Insurance Co. v. California, 1509 U.S. 764 (1993), and EU judgments on extraterritoriality); *see also* Giorgio Monti, *The Global Reach of EU Competition Law*, *in* EU LAW BEYOND EU BORDERS: THE EXTRATERRITORIAL REACH OF EU LAW (Marise Cremona & Joanne Scott eds., 2019).

415. F. Hoffman-La Roche Ltd. v. Empagran S.A., 542 U.S. 155, 165 (2004). *But see* Ralf Michaels, *Supplanting Foreign Antitrust*, 79 L. & CONTEMP. PROBS. 223, 223–47 (2016)

competition authorities were to police anticompetitive conduct abroad, conventional economic analysis would be focused on neatly-delineated domestic markets. All of this might seem unremarkable and rather "normal," but it is increasingly out-of-sync with the reality of global data generation and transnationally distributed yet interconnected data infrastructures. As both Tim Mitchell and Hugo Radice have shown, the idea of a national economy that is congruent with the nation state is a construct whose creation can be attributed to (perceived) econometric necessities.[416] Most economists' traditional focus on national markets corresponds to most lawyers' traditional focus on national law. This framing leads to a misalignment between economic and legal frameworks and a data reality where conventional territorial borders are far less relevant to delineating the components and dimensions of data infrastructures. International law or some other form of inter-public or even "global" law could potentially be used to remedy such misalignment.[417] The competition chapters in contemporary trade agreements achieve nothing to that effect, as they remain mainly concerned with procedural rights of businesses during investigations or push back against governmental interference in markets, while ignoring the possibility that globally-distributed but centrally-controlled market power may be more than the sum of its parts.[418] Indeed, antitrust and competition law are not concerned with the global concentration of market power. Thus, they are structurally ill-equipped to confront data inequality arising from global control over data infrastructures nurtured by such market power.

III. CONFRONTING DATA INEQUALITY FOR DIGITAL DEVELOPMENT

The dominant narrative in the development discourse tends to emphasize the welfare gains stemming from digitalization, data sharing, and data-driven technologies (particularly artificial

(arguing that developed country with effective antitrust enforcement should lend their antitrust enforcement capacity to developing countries).

416. *See generally* Tim Mitchell, *Origins and Limits of the Modern Idea of the Economy* (Advanced Study Ctr., Univ. Michigan, Working Papers No. 12, 1995); Hugo Radice, *The National Economy: A Keynesian Myth?*, CAP. & CLASS, Spring 1984, at 111.

417. On international law as inter-public law, see generally Benedict Kingsbury, *International Law as Inter-Public Law*, 49 NOMOS 167 (2009). For a discussion of varieties, possibilities, and limitations of "global law," see generally Neil Walker, INTIMATIONS OF GLOBAL LAW (2014). On Global Data Law, see generally Global Data Law, *supra* note 6.

418. *E.g.*, USMCA, *supra* note 276, art. 21.2.

intelligence/machine learning).[419] International organizations devoted to the promotion of economic growth and human flourishing urge the expanded use of digital data for economic gain and social benefit. Although they increasingly consider the drawbacks of digitalization and the risks associated with data, the focus is often on adverse effects on individual privacy (and occasionally other) rights and on security implications caused by increasing reliance on interconnected systems.[420] Efforts to promote good "data governance" practices often seek to maximize the value of data as an asset, target predominantly public sector actors, and draw inspiration from the extant legal frameworks that we discussed in Part II.[421]

While we do not discount the importance of these efforts, we want to draw attention to a complementary need to consider the unequal power to determine *what* becomes data and, conversely, what does *not* become data. This perspective requires an increased consideration of the power dynamics inherent in the social practices through which data is generated. In this context, we have also highlighted the role of infrastructural control in constituting and entrenching unequal control over data, particularly when centralized in the hands of corporate actors.[422] We have examined the limitations of extant dominant legal approaches in remedying data inequality as well as the risk that they may, in certain circumstances, further entrench data inequality.[423]

In this final Part, we address some interventions that might aid in remedying data inequality, with particular attention to development freedoms of individuals and communities. While we focus on countries with developing digital economies, we emphatically do not believe that data inequality can be resolved in a wholesale fashion without due regard to the particular economic, social, cultural, and

419. *See, e.g.*, WORLD BANK, WORLD DEVELOPMENT REPORT 2021: DATA FOR BETTER LIVES 93 (2021), https://www.worldbank.org/en/publication/wdr2021 [https://perma.cc/JQW9-DKNV].

420. *Id*. ch. 6 (discussing safeguards necessary for deriving value from data for development).

421. *See, e.g.*, OECD, THE PATH TO BECOMING A DATA-DRIVEN PUBLIC SECTOR 25 (2019) ("Good data governance is imperative for governments that aim to become more data driven as part of their digital strategy. It can help to extract value from data assets, enabling greater data access, sharing and integration at the organizational level and beyond, and increasing overall efficiency and accountability."); *see also* Arturo Muente-Kunigami, *We Need to Urgently Review Our Data Governance Frameworks*, OPEN DATA CHARTER (2020), https://opendatacharter.medium.com/we-need-to-urgently-review-our-data-governance-frameworks-202a13142957 [https://perma.cc/S5KJ-WG6X].

422. *See supra* Part I.

423. *See supra* Part II.

historical contexts in which different countries and their populations find themselves.[424] To the contrary, it follows naturally from our prior discussion of how data infrastructures shape how and what data is collected and used, and for what purpose that contextualization and self-determined experimentation are important pathways towards more equitable and democratic digital development.

At present, it appears that countries looking to develop digital economies are faced with a dilemma. On the one hand, a lack of prerequisite physical or digital components (and the high costs associated with building them domestically) may lead countries to rely on infrastructure provided by the world's leading tech companies, which are overwhelmingly based in the United States and China. This approach is encouraged by the logic of market efficiencies. Most governments and international organizations proceed on the assumption that capitalism produces economically superior outcomes (at least in the aggregate).[425] On the other hand, because data infrastructures *shape* data and consequently the representations of physical, social, or political phenomena that data aims to capture and reflect (albeit imperfectly), the interest of public constituencies may pull towards a more local and collectivist control over data infrastructures and the development of the necessary technical, social, and organizational structures and practices. These latter interests are being supported by growing doubts about, and occasional resistance to, an unconditional embrace of a data-driven capitalist economic development model, prompted by the excesses of "surveillance capitalism," the dominance of platforms in "informational capitalism," and the post-colonial continuities of data extractivism and exploitation.[426] Some scholars even wonder whether some form of "digital socialism" might be economically viable after all, contrary to

---

424. David M. Trubek, *Law and Development: Forty Years After 'Scholars in Self-Estrangement'*, 66 UNIV. TORONTO L.J. 301, 318 (2016):

> Three important ideas have helped shape twenty-first century law and development: the understanding that development does not follow prescribed script but requires constant experimentation; the recognition that capitalism can take many forms and law will vary with the dominant form of market system; and the idea that legal rights are part of what is meant by development, not just a means to an end.

425. *See* BRANCO MILANOVIC, CAPITALISM, ALONE: THE FUTURE OF THE SYSTEM THAT RULES THE WORLD ch. 1 (2019) (observing that there is no longer a contest between different economic systems, but rather a question of which variety of capitalism to embrace, and how much state involvement to allow or require). This does not mean, however, that alternatives to capitalism are not imaginable or achievable. *See generally* ERIK OLIN WRIGHT, ENVISIONING REAL UTOPIAS (2010).

426. *See generally* ZUBOFF, *supra* note 33; COHEN, *supra* note 14. *See also* Couldry & Mejias, *supra* note 95, pt. I.

Friedrich Hayek's assumptions.[427] What if the unprecedented generation of data—though, thus far, highly concentrated in the hands of few—and the resulting ability to generate information, make a collectively governed economy and society plausible, or even desirable?[428] The success of entities that rely on data-dependent and highly-centralized planning for their commercial success lends credence to this possibility.[429]

The need for economies and societies around the world to shape their individual and collective digital destinies coincides with, and corresponds to, renewed calls to question long-standing tenets of the development agenda. The traditional critique of the development agenda's emphasis on economic growth, conventionally measured in the aggregate, is now being reinforced by those who question the role of information and communication technologies in development and who advocate for a shift towards other paramount values and objectives, such as human dignity or individual and collective freedom.[430] As the world confronts a global climate crisis, the environmental cost of increased digitalization ought to be (re)considered.[431] The intensifying debate about economic inequality must now also confront the impact of digitalization, in particular data inequality-induced information asymmetries and "winner-takes-all" dynamics common in Western digital economies and even more

427. Evgeny Morozov, *Digital Socialism? The Calculation Debate in the Age of Big Data*, 116/117 NEW LEFT REV. 33, 36 (2019) (in response to Viktor Mayer Schönberger & Thomas Ramge, REINVENTING CAPITALISM IN THE AGE OF BIG DATA (2018)).

428. Przemysław Pałka, *Algorithmic Central Planning: Between Efficiency and Freedom*, 83 L. & CONTEMP. PROBS. 145, 146 (2020).

429. LEIGH PHILLIPS & MICHAEL ROZWORSKI, THE PEOPLE'S REPUBLIC OF WALMART: HOW THE WORLD'S BIGGEST CORPORATIONS ARE LAYING THE FOUNDATION FOR SOCIALISM 212–13 (2019).

430. *See, e.g.*, Manuel Castells & Pekka Himanen, *Introduction*, *in* RECONCEPTUALIZING DEVELOPMENT IN THE GLOBAL INFORMATION AGE 1 (Manuel Castells & Pekka Himanen eds., 2014) (arguing to "reconceptualize human development as the fulfilment of human wellbeing in the multidimensionality of the human experience, ultimately affirming dignity as the supreme value of development"). Amartya Sen's capabilities approach is often at the heart of such endeavors. *See generally* Devinder Thapa & Øystein Sæbø, *Exploring the Link between ICT and Development in the Context of Developing Countries: A Literature Review*, 64 ELEC. J. INFO. SYS. DEV. COUNTRIES 1 (2014); Richard Heeks & Jaco Renken, *Data Justice for Development: What Would It Mean?*, 34 INFO. DEV. 90 (2016).

431. For a discussion of environmental harms caused by cloud computing and processing of very large datasets for machine learning, see generally Elettra Bietti & Roxana Vatanparast, *Data Waste*, 61 HARV. INT'L L.J. FRONTIERS (2020). *See also* Emma Strubell, Ananya Ganesh & Andrew McCallum, *Energy and Policy Considerations for Deep Learning in NLP*,PROC.OF THE 57TH ANNUAL MEETING OF THE ASSOC. FOR COMPUTATIONAL LINGUISTICS,3645, 3645 (2019), https://arxiv.org/pdf/1906.02243v1.pdf [https://perma.cc/D3NT-PQDN].

pronounced transnationally.[432] The COVID-19 pandemic accelerated the embrace of certain affordances provided by digital technologies, most notably internet-enabled video calls. The pandemic also called into question established development priorities, with considerably more emphasis being placed on wellbeing, income equality, and environmental sustainability.[433]

We cannot do justice to these broader questions and debates in this Article, but we seek to acknowledge their existence in order to situate the interventions that we suggest and to recognize their limitations. Developing economies may be wise to retain the freedom to experiment with different developmental approaches, given the lack of robust empirical evidence about optimal approaches to digital development and given the salience, longevity, and path dependencies of infrastructures.

Addressing data inequality inevitably requires confronting competing values and interests. As we have emphasized throughout this Article, control over data and the power to datafy arise from control over data infrastructures. These infrastructures are neither agentless nor static; rather, they are deeply political and dynamic. Redistributing existing data (e.g., by making data "open" or by encouraging data philanthropy), or moving away from concentrated infrastructural control to more distributed configurations, does not necessarily equalize the power to determine what should and should not be datafied.[434] Focusing solely on questions about how data should be generated (and by whom), without attention to how concentrations of data are accumulated via infrastructural control and how such control is established and maintained, including by exploiting legal ambiguities, risks missing relevant sites for intervention.

---

432. As Dan Ciuriak explained, "the business model of the data-driven economy is based on exploitation of information asymmetry. By further extension, there are fundamental information asymmetries between countries that can build companies on data assets and those that cannot. Information asymmetry is, in some sense, the 'original sin' of the data-driven economy." DAN CIURIAK, RETHINKING INDUSTRIAL POLICY FOR THE DATA-DRIVEN ECONOMY 6 (2018); *see also* ERIK BRYNJOLFSSON & ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES 154–55 (2014).

433. *See* Kathleen R. McNamara & Abraham L. Newman, *The Big Reveal: COVID-19 and Globalization's Great Transformations*, 74 INT'L ORG. E59, E73 (Supp. 2020) (assessing the impact of COVID-19 on globalization and asserting that the pandemic has underscored the importance of digital technologies).

434. Indeed, in some instances, concentration of infrastructural control might be necessary to empower a previously disempowered constituency. Illustrative of this is the case of indigenous peoples' struggles to reclaim the power to create knowledge about themselves, leading to claims of indigenous data sovereignty. *See* Pool, *supra* note 43, at 69–71.

Data inequality needs to be addressed in tandem with the contemporary capitalist logics and dynamics within which it is situated.[435] An effective reshaping of these logics and dynamics may require a reshaping of the law as well. For such an endeavor to succeed, lawyers ought to critically scrutinize the many ways in which law, including international law, has contributed to inequality past and present.[436] Legal doctrines, regulatory dogmas, and technical specifications are not going to remedy data inequality unless such interventions are also accompanied by continuous, iterative, and inclusive public debate, as well as deliberation, contestation, decision-making, and implementation. We thus caution development aid agencies and other similarly positioned actors not to close off, but instead to encourage, the expansion of spaces that would allow for such activities to materialize in developing as well as developed economies.[437]

With that preface, in the following Sections, we advance several ideas that might chart pathways for confronting or mitigating data inequality. Some of these ideas may, at first blush, seem experimental and radical, while others may appear to be more iterative and pedestrian. Given the complex interaction of technical, social, and organizational dynamics in data infrastructures, interventions in any one of these dimensions will inevitably produce ripples in others. To that end, we propose both interventions that have direct regulatory effects and others that create enabling environments for more meaningful political consideration and the contestation of datafication and digital development. We progress from large-scale and

435. *See generally* COHEN, *supra* note 14 (analyzing how law enables informational capitalism); PISTOR, *supra* note 10, at 167–68, 224 (emphasizing the role of lawyers in coding capital and shielding it from the democratic process); Kapczynski, *supra* note 14, at 1496–1514 (showing how law protects certain data related interests of informational capitalists but not others through mechanisms of "encasement"). *See generally* Viljoen, *supra* note 287.

436. There is extensive scholarship in critical legal studies on these questions. For seminal work on intersectionalist perspectives on law, see generally Kimberlé Crenshaw, *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine*, U. CHI. LEGAL F. 139 (1989). For critical analysis of international law, see generally NTINA TZOUVALA, CAPITALISM AS CIVILISATION: A HISTORY OF INTERNATIONAL LAW (2020); Antony Anghie et al. (eds.), THE THIRD WORLD AND INTERNATIONAL ORDER: LAW, POLITICS AND GLOBALIZATION (2004); James Thuo Gathii, *The Promise of International Law: A Third World View*, GROTIUS LECTURE AT THE 2020 VIRTUAL ANNUAL MEETING OF THE AMERICAN SOCIETY OF INTERNATIONAL LAW (June 25, 2020). *See also infra* Section III.A. (discussing international economic law).

437. *See generally* Amba Kak, *"The Global South Is Everywhere, But Also Always Somewhere": National Policy Narratives and AI Justice*, AIES '20: PROCEEDINGS OF THE AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY (Feb. 2020), https://doi.org/10.1145/3375627.3375859 [https://perma.cc/B2FK-VKYK].

foundational recommendations to more local and targeted suggestions. We consider different actors and scales in the ensuing analysis. It cannot, nor should it, be assumed that nation states, and their peoples and territories, are necessarily the only or most suited actors and scales to address data inequality, which is, fundamentally, both a global and local phenomenon.[438]

Data infrastructures and the entities that control them transgress territorial borders with relative ease, while the interests of affected publics might be transnationally aligned or in tension with one another.[439] At the same time, nation states remain the dominant form of organized political power, and maybe even more so in a world with stark anti-globalist currents. Moreover, they remain the main subjects and objects of international law, as traditionally conceived, and are tasked with steering economic development within the framework of the global economic order as currently constituted. For this reason, we direct our attention mainly towards governments and international organizations.

Our large-scale recommendations can be broadly summarized as: (1) encouraging the retention of development freedom; (2) reclaiming infrastructural control; (3) demanding transparency; (4) pooling and differentiating access to data; and (5) developing collective data governance mechanisms. In the following Sections, we consider each of these ideas, highlighting their respective legal and infrastructural elements alongside their relevance for confronting data inequality. Their realization and success will, of course, depend on support by relevant political actors and social movements. We cannot offer an account of how the necessary level of such support could be generated, but we nonetheless believe and hope that laying out ideas about what could be done will at least affect the discourse around data inequality and digital development more generally and might spur relevant actors into action.

A. Retaining Developmental Freedom

Data is growing in importance as a medium for economic, social, and political ordering and as a resource for economic development. Thus, governments are tasked with managing the

438. *See* YANNI A. LOUKISSAS, ALL DATA ARE LOCAL: THINKING CRITICALLY IN A DATA-DRIVEN SOCIETY 22 (2019) ("[D)ata may be shaped by local conditions, yet they serve a combination of needs, near and far . . . . [T]here is no global experience of data, only an expanding variety of local encounters.").

439. On the interplay between infrastructures, publics and law, see generally Kingsbury & Maisley, *supra* note 84.

transition towards increasingly digitally mediated economies and societies. As we have seen, their choices may be constrained by existing and emerging commitments under international economic law which may contribute to data inequality.[440] New agreements in the mold of the USMCA and the USJDTA favor global access to governmental data by encouraging open data policies, while limiting states' abilities to impose restrictions on cross-border data transfers or to mandate the use of domestic computing facilities.[441] These agreements seek to carry forward core tenets of the world's trading system, as constituted under the WTO, by applying and extending policy prescriptions and economic development theories to an increasingly data-driven economy.[442] Despite withdrawing from the TPP itself, the United States successfully inserted its favored rules into instruments of international economic law that are now being advanced and amplified by other countries.[443] The reason for this outcome is likely the belief in a certain model of digital development that can be termed "Silicon Valley Consensus." The Silicon Valley Consensus emphasizes unrestricted data flows, pushes back against "data localization," requires only a mere minimum of data protection regulation, and is generally predisposed against state intervention in the digital economy.[444] States that embrace this model commit to "free" data flows as a mode for digital development not just vis-à-vis

440. *See supra* Sections II.A–D.

441. Streinz, *supra* note 199, at 188. Open data is discussed in more detail *infra* Section III.D.

442. *See supra* note 276 and accompanying text; Agreement Between the United States of America and Japan Concerning Digital Trade, Japan-U.S., Oct. 7, 2019, T.I.A.S. No. 20-0101.1. While potentially effective rhetorically and ideologically, it is conceptually unconvincing to analogize conventional trade in goods and services to "digital trade" just because data "flows" across borders. The most recent agreements—such as the Digital Economy Partnership Agreement pioneered by Singapore, Chile, and New Zealand, and the Digital Economy Agreement between Australia and Singapore—drop the "trade" moniker and only speak of "digital economy agreements," while retaining the rules and substantive concepts that the United States pioneered through TPP, USMCA, and USJDTA. *See generally* Digital Economy Partnership Agreement, Chile-N.Z.-Sing., June 11, 2020, https://www.mfat.govt.nz/assets/Trade-agreements/DEPA/DEPA-Signing-Text-11-June-2020-GMT-v3.pdf [https://perma.cc/Y64K-49AX].

443. *See* Shamel Azmeh & Christopher Foster, *The TPP and the Digital Trade Agenda: Digital Industrial Policy and Silicon Valley's Influence on New Trade Agreements* 5 (NTALSE Int'l Dev. Working Paper Series No. 16-175 2016), https://www.lse.ac.uk/international-development/Assets/Documents/PDFs/Working-Papers/WP175.pdf [https://perma.cc/Q4PV-3LPT]. *See also* Streinz, *supra* note 136, at 314–15.

444. Streinz, *supra* note 136, at 330. *See generally* Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639 (2014).

each other. Multinational corporations are easily able to avail themselves of the prerequisite corporate nationality to invoke treaty protections against regulatory measures that limit cross-border data transfers or require the use of domestic computing facilities.[445]

The European Union has realized that this model is at odds with its data protection regime, which limits cross-border data transfers according to the (perceived) level of data protection afforded in other jurisdictions.[446] Accordingly, the European Union is advancing commitments under international economic law that echo the European conceptualization of data protection as a fundamental right and that shield the GDPR from scrutiny and anti-regulatory pressure under international economic law.[447] In other respects, however, the stance of the European Union is aligned with the Silicon Valley Consensus in pushing back against those modes of data localization in which the European Union itself is not engaged (e.g., requirements for local storage). China, on the other hand, has largely refrained from advancing its policy preferences for data governance through instruments of international economic law.[448] China is, however, affecting data governance beyond its borders in other ways, including through digital infrastructure investments; in November 2020, China was one of fifteen countries in the Asia-Pacific that signed the Regional Comprehensive Economic Partnership (RCEP) Agreement, which contains an e-commerce chapter modeled after the TPP but with considerably more leeway for governments to retain restrictive policies.[449] Even with these significant carveouts in place, India,

445. Streinz, *supra* note 10.

446. GDPR, *supra* note 184, art. 45.

447. Horizontal Provisions, *supra* note 302, art. A(1).

448. *See* Erie & Streinz, *supra* note 166, at 21–23.

449. *Compare* Regional Comprehensive Economic Partnership Agreement, arts. 12.14, 12.15, Nov. 15, 2020, https://rcepsec.org/wp-content/uploads/2020/11/All-Chapters.pdf [https://perma.cc/32YG-DLVE], *with* Trans-Pacific Partnership Agreement, arts. 14.11, 14.13, Feb. 4, 2016, https://ustr.gov/sites/default/files/TPP-Final-Text-Electronic-Commerce.pdf [https://perma.cc/F2FQ-A9ML]. RCEP footnotes 12 and 14 make clear that it is for the implementing party to decide—and not for other parties or a dispute settlement body to second guess—whether a measure is "necessary." *See* Thomas Streinz, *RCEP's Contribution to Global Data Governance*, AFROMONICSLAW, at 3 (Feb. 19, 2021), https://www.afronomicslaw.org/category/analysis/rceps-contribution-global-data-governance-0 [https://perma.cc/ACD5-P2Y9]. The RCEP is designed to intensify economic ties between the ten ASEAN members and the non-ASEAN countries of China, South Korea, Japan, Australia, and New Zealand. *See generally* Pasha L. Hsieh, *The RCEP, New Asian Regionalism and the Global South* (N.Y.U. Inst. For Int'l L. & J., Working Paper 2017/4, 2017), https://www.iilj.org/wp-content/uploads/2017/12/Hsieh-IILJ_2017_4_MegaReg.pdf [https://perma.cc/2PME-LYUK].

which had been part of the negotiations earlier, refrained from signing the RCEP. In September 2021, China formally requested accession to the CPTPP, but it remains to be seen whether formal negotiations will ensue.[450] If so, China will likely push for accommodations for its data governance framework along the lines of those included in the RCEP.

Developing economies are confronted with three models for new data-related commitments under international economic law: (1) they can follow the Silicon Valley Consensus as instantiated in the TPP and carried forward in further agreements, including the Digital Economy Partnership Agreement (DEPA);[451] (2) they may seek alignment with the European Union's pro-regulatory position centered on data protection values; or (3) they may favor RCEP's new model, which grants broad and self-judging exceptions. The alternative is to refrain from entering into new commitments altogether, as exemplified by India's stance in the WTO and elsewhere.[452] In considering these options, it is worth noting that the global digital economy developed without much support from international economic law. The attempts to create new rules in instruments of international economic law, as instantiated by the Silicon Valley Consensus, are more about *entrenching* a vision of the digital economy without forceful regulatory intervention at a moment at which such interventions are on the rise. The data inequality dimensions that we have identified in this Article have been largely ignored in the conventional discourse around "digital trade" and "electronic commerce." This is, in part, a continued legacy of the embedded liberalism that carried the world trading system after World War II and relegated distributional questions to

450. Wendy Cutler, *A Trade Pact Faces a Crucial Test as China and Taiwan Try to Join*, BARRON'S (Oct. 27, 2021, 5:00 AM), https://www.barrons.com/articles/a-trade-pact-faces-a-crucial-test-as-china-and-taiwan-try-to-join-51635282396 [https://perma.cc/BW8J-KFTN].

451. The DEPA was signed (electronically) by Chile, New Zealand, and Singapore in June 2020 and has been in force between Singapore and New Zealand since December 2020. The DEPA builds on the TPP, which originated initially in an agreement between the same countries. But the DEPA goes significantly further by creating new provisions hitherto not seen in international economic agreements, openly announcing itself as a digital economy agreement rather than an agreement merely on "electronic commerce" or "digital trade." *See* Digital Economy Partnership Agreement, *supra* note 442.

452. The economic policy calculation is complicated by a pervasive and somewhat paradoxical lack of data about the digital economy, with neither conventional economic nor trade statistics accounting sufficiently for the value of data and the significance of data flows. It will ultimately depend on countries' particular economic context and trajectory whether or not signing on to any of the currently available models is advisable or whether refraining from entering into such, potentially long lasting, commitments is the more prudent course of action.

states' domestic social welfare systems.[453] It also reflects, however, an economistic conceptualization of data as a rent-generating asset which tends to conceal other dimensions of data inequality.[454] Imagining and designing instruments of international economic law that are more attuned to dimensions of data inequality strikes us as a promising yet uncharted path forward.[455]

Addressing inequalities of the digital economy thus requires novel interventions that may conflict with the theories and concepts under which the world trading system has operated since World War II. This reality has institutional implications for the WTO, which sits at the heart of this system, as well as for development organizations more broadly as they advise countries on which policies to pursue. It is in this context that we caution against premature commitments and suggest that retaining the ability to experiment with different development strategies, digital industrial policies, and attendant social policies might be the more prudent course of action to confront data inequality going forward.[456]

For the moment, some developing economies have pushed back against new commitments on data governance in the WTO's work program on electronic commerce, which is now proceeding as a plurilateral initiative.[457] The drafters of the African Continental Free

453. *See* Andrew T. F. Lang, *Reconstructing Embedded Liberalism: John Gerard Ruggie and Constructivist Approaches to the Study of the International Trade Regime*, 9 J. INT'L ECON. L. 81, 97–98 (2006). Joseph Stiglitz's GLOBALIZATION AND ITS DISCONTENTS 16 (2003) and Dani Rodrik's THE GLOBALIZATION PARADOX: DEMOCRACY AND THE FUTURE OF THE WORLD ECONOMY (2011) are prominent articulations of critiques of the world trading system and its impact on social welfare. *See generally* Sonia E. Roland & David Trubek, *Embedded Neoliberalism and Its Discontents: The Uncertain Future of Trade and Investment Law*, *in* WORLD TRADE AND INVESTMENT LAW REIMAGINED 87 (Alvaro Santos, Chantal Thomas & David M. Trubek eds., 2019).

454. *See supra* Sections I.A–B.

455. *See* Kingsbury et al., *supra* note 135, at 60 (imagining truly 21st-century agreements, potentially based on the welfare conception championed by Amartya Sen).

456. *See generally* Dan Ciuriak, *Digital Trade: Is Data Treaty-Ready?*, CIGI PAPER NO. 162 (Feb. 21, 2018), https://www.cigionline.org/publications/digital-trade-data-treaty-ready [https://perma.cc/YY8G-2M6A].

457. Seventy-one WTO members signed a joint statement on electronic commerce during the WTO's 11th Ministerial Conference in Buenos Aires. World Trade Organization, Ministerial Joint Statement on Electronic Commerce, WTO Doc. WT/MIN(17)/60 2017). Negotiations between seventy-six WTO members commenced in January 2019. World Trade Organization, Joint Statement on Electronic Commerce, WTO Doc. WT/L/1056 2019). As of January 2020, eight-three WTO members were participating in the negotiations, including all developed countries, but only five WTO members from Africa (Benin, Nigeria, Cote d'Ivoire, Kenya, and Cameroon), and no Caribbean or developing Pacific Island countries. *See* YASMIN

Trade Area (AfCTA) agreement did not include provisions on data governance.[458] At the same time, states' existing commitments under international investment law, enshrined in investment chapters of trade agreements and bilateral investment treaties, may be the more consequential constraints in the short-term, especially if states resort to mandatory data-sharing requirements to redistribute data.[459] International investment law does not just have anti-regulatory effects on public law measures, but it may also shape and reshape core concepts of private law.[460] It is likely to be mobilized to contest governmental regulation in the digital domain and may also influence the evolving debate around legal rights to data and questions of legal data ownership.[461] Countries entered into these commitments when digitalization was not yet on the horizon or was not as pressing as it is now. Resisting the mobilization of international investment law to protect existing highly asymmetric and concentrated control over data and data infrastructures will be an important challenge for development lawyers going forward and will require prudent judgment by the arbitrators who will be tasked with resolving these conflicts.[462]

States may have more leeway in challenging unequal control over data and data infrastructures if they are relatively unconstrained from existing commitments under international economic law in terms of cross-border data transfers and protections of data as an asset under international investment law. They could, for example, adopt regulatory frameworks that specifically target infrastructural control by platform companies and, where appropriate, mandate access to data

---

ISMAIL, INT'L INST. FOR SUSTAINABLE DEV., E-COMMERCE IN THE WORLD TRADE ORGANIZATION: HISTORY AND LATEST DEVELOPMENTS IN THE NEGOTIATIONS UNDER THE JOINT STATEMENT 14 (2020), https://www.iisd.org/publications/e-commerce-world-trade-organization-history-and-latest-developments-negotiations-under [https://perma.cc/V776-L2E5]. In February 2021, India and South Africa formally criticized these initiatives as in tension with WTO principles of consensus-based multilateralism. World Trade Organization, The Legal Status of 'Joint Statement Initiatives' and Their Negotiated Outcomes, WTO Doc. WT/GC/W/819, at 2, (2021).

458. *See* Chijioke Chijioke-Oforji, *The Untapped Potential of the African Continental Free Trade Agreement in the African E-commerce Agenda*, 27 INT'L T.L.R. 141, 151–53 (2021) (arguing for the inclusion of e-commerce provisions in future negotiations).

459. *See* Streinz, *supra* note 199, at 182.

460. *See generally* Arato, *supra* note 293.

461. COHEN, *supra* note 14, at 257–60; *see also supra* Section II.B.

462. Where confidence in their ability to resolve such disputes in an equitable and just ways is lacking, withdrawal from the international investment system as currently constituted might be worth exploring. Digital development in a data-driven economy might follow a different logic than in the knowledge-based economy which relied on cheap manufacturing and far-flung global value chains. *See* CIURIAK, *supra* note 432, at 14.

for public and commercial actors. Additionally, and especially where regulatory power cannot be effectively asserted, states could pursue and support the development of independent data collection capacity and commensurate data infrastructures. Indeed, if financially and technologically feasible, the latter route might be superior to outright data sharing, as the relevant publics could determine for themselves which data ought to be collected and how, instead of those choices being dictated by other data collectors in pursuit of their own interests.[463] These options are discussed in the following Section.

B. Reclaiming Infrastructural Control

Much of this Article has focused on the concentration of infrastructural control in the hands of corporate actors.[464] Although predominantly based in the United States and China, and to a far lesser extent in Europe, many large tech companies have been entering previously untapped markets, notably developing economies in Africa and Asia.[465] Often this kind of corporate expansion is being encouraged and welcomed in the hope that increasing the supply of resources, expertise, and access to digital technologies might enable developing economies to "leapfrog" in their economic development.[466] Without passing judgment on the merits of this proposition, too often sincere desire for efficiency and quick returns fails to take full account of the associated financial, social, and political costs, and glosses over the long-term sustainability of infrastructural dependencies. To be clear, we are not advocating for digital development in isolation nor are we suggesting that developing economies should forgo any or all services provided by platform companies, cloud, or other data

463. We stress that the normative desirability of these interventions is contingent on the democratic credentials and public values of the political systems that bring them about. If a democratic polity decides to condition or otherwise restrict cross-border data transfers to prevent data extraction, the normative evaluation ought to be different compared to an autocratic imposition of data transfer limitations for reasons of authoritarian self-preservation.

464. *See supra* Section I.C.

465. *See, e.g.*, Annie Njanja, *Google Confirms $1B Investment Into Africa, Including Subsea Cable for Faster Internet*, TECHCRUNCH (Oct. 6, 2021, 5:10 AM), https://techcrunch.com/2021/10/06/google-confirms-1b-investment-into-africa-including-subsea-cable-for-faster-internet [https://perma.cc/BM52-NP2U]; Laura Grunberg, *China's Tech Companies Are Going Global – and Remaking China's Image in the Process*, DIPLOMAT (June 4, 2021), https://thediplomat.com/2021/06/chinas-tech-companies-are-going-global-and-remaking-chinas-image-in-the-process/ [https://perma.cc/EG8Y-ZF2C].

466. *See, e.g.*, Makhtar Diop, *Africa Can Enjoy Leapfrog Development*, WORLD BANK OP. (Oct. 11, 2017), https://www.worldbank.org/en/news/opinion/2017/10/11/africa-can-enjoy-leapfrog-development [https://perma.cc/7LJE-9H4A].

infrastructure providers. Rather, we caution against schematic efficiency and necessity narratives as default positions to justify the privatization and corporatization of public services. Concentrated corporate control over critical data infrastructures should not be a quasi-automatic default position. Instead, we advocate for considerate and creative data infrastructure planning that engages the relevant publics, who ought to decide for themselves how their environments and lives are to be or not to be datafied.[467]

In this Section, we consider regulatory options that depend on considerable state power. For this reason, we emphasize again that concentrated infrastructural control over data in the hands of governments can also be cause for concern. The regulatory frameworks that we discuss position individual states against foreign data-infrastructure-controlling corporations, but it is conceivable and perhaps even quite likely that new data jurisdictions will emerge—connecting different publics, creating transnational alliances, and presenting different regulatory options that are not aligned with jurisdictional control of a single state.[468] We allude to such possibilities in more detail in our discussion of collective data governance in Section III.E. Here, we highlight the potential role that international organizations might play in creating and mediating such new data jurisdictions.

Reclaiming infrastructural control can take different forms. Recent EU initiatives, discussed in Part II above, have opened the door to a regulatory approach that specifically targets online platform companies, but do so in a differentiated manner, imposing additional obligations on platforms of particularly large reach.[469] This may be a promising regulatory pathway for countries seeking to benefit from resources and infrastructures offered by dominant actors while preserving (or fostering) public control over data infrastructures. Although a direct "transplant" of EU law to other countries would be counterproductive—given the European Union's unique political and economic specificities—useful lessons nonetheless can be gleaned from the EU regulatory agenda. The proposed DSA and DMA seek to reassert public authority over corporate actors with concentrated

467. On reinvigorating planning and foresight in the context of infrastructures, see generally Benedict Kingsbury, *Infrastructure and InfaReg: On Rousing the International Law 'Wizards of Is*,*'* 8 CAMBRIDGE INT'L L.J. 171 (2019).

468. *See* Marietje Schaake & Tyson Barker, *Democratic Source Code for a New U.S.-EU Tech Alliance*, LAWFARE (Nov. 24, 2020, 3:07 PM), https://www.lawfareblog.com/democratic-source-code-new-us-eu-tech-alliance [https://perma.cc/4P72-WP3Z] (calling for a transatlantic alliance against big tech).

469. *See supra* Section II.D.

power over infrastructures and the power to datafy. These proposals might indicate a shift away from the prevalent tendency to treat data as a regulatory object towards a regulation of infrastructural control that transcends established concepts under competition law.[470]

It may be difficult for smaller states, without market or political power comparable to the European Union, to assert regulatory control over large U.S.- and China-dominated tech companies, or to negotiate their own terms.[471] In addition to traditional lobbying activities, multinational corporations may invoke commitments under international economic law to thwart regulatory initiatives.[472] Indeed, governments and corporations may leverage infrastructural control itself to push against regulatory efforts by smaller states. Henry Farrell and Abe Newman have drawn attention to the dependencies created by inter-networked technologies.[473] States with control over the companies that build, operate, and maintain these infrastructures gain widespread access to data (through surveillance measures) and may mobilize their control over chokepoints (e.g., by threatening cut offs) to advance their geopolitical objectives.[474] The resulting power differentials not only shape geopolitical confrontation and alignment but also affect economic development pathways, particularly for developing economies.[475] China promotes digital infrastructure investments through its Digital Silk Road, which forms part of the larger Belt and Road Initiative and promises "data sovereignty." Such promise, however, is tenuous, as Chinese technology companies acquire central roles within the relevant data infrastructures transnationally and may attain control over thus-generated data.[476] Over the course of 2020, India repeatedly took the extraordinary step of banning outright certain Chinese apps, citing concerns over the

470. *See supra* Section II.D.

471. Indeed, it is not yet known whether the European Union's regulatory efforts will succeed in curbing the infrastructural power of platform companies.

472. *See* Tim Dorlach & Paul Mertenskötter, *Interpreters of International Economic Law: Corporations and Bureaucrats in Contest over Chile's Nutrition Label*, 54 L. & SOC'Y REV. 571, 586–93 (2020) (tracing how the transnational food industry challenged food labelling regulation in Chile).

473. Henry Farrell & Abraham L. Newman, *Weaponized Interdependence*: *How Global Economic Networks Shape State Coercion*, 44 INT'L SEC. 42, 45 (2019).

474. *Id*. at 65. *See generally* Cartwright, *supra* note 150.

475. Amrita Narlikar, *Must the Weak Suffer What They Must? The Global South in a World of Weaponized Interdependence*, *in* THE USES AND ABUSES OF WEAPONIZED INTERDEPENDENCE 289, 290 (Daniel W. Drezner, Henry Farrell & Abraham L. Newman eds., 2021).

476. Erie & Streinz, *supra* note 166, at 86–87.

"sovereignty and integrity of India."[477] Western platform companies have not yet faced such prohibitions, but when subjected to increased regulatory scrutiny or taxation demands, they have repeatedly threatened to withdraw their services.[478] While such threats may seem like a mere negotiation tactic to influence lawmakers and the public, they may also constitute a leveraging of infrastructural control, the effectiveness of which will depend on the interdependencies and scale commanded by the relevant platforms.

Another approach to resisting the dominance of U.S.- and China-based companies is to invest in the development of alternative public (or public-private) data infrastructures. Even prior to its recent regulatory initiatives, the European Union supported the development of a European, independent cloud infrastructure called GAIA-X and promised to make this infrastructure available to others, conditional on adherence to European law, in particular the GDPR.[479] While the technological and economic success of this initiative remains to be seen, and its appeal to entities outside the European Union remains uncertain, GAIA-X signals an increasing awareness of infrastructural dependencies amidst rising geopolitical contestation between the United States and China, which are home to the world's leading cloud

477. *India Bans 43 More Mobile Apps as It Takes on China*, REUTERS (Nov. 25, 2020, 12:35 AM), https://www.reuters.com/article/india-china-apps-idINKBN2850I2 [https://perma.cc/K249-MTD8].

478. Some illustrative examples include the following: (1) In light of the ECJ's decision in *Schrems II* (see *supra* Section II.A, particularly note 186 and accompanying text), Facebook declared vis-à-vis the Irish High Court that, in the event of a complete prohibition on the transfer of user data to the United States, it was not clear how Facebook could continue to provide Facebook and Instagram services within the European Union. See the sworn affidavit by Facebook's head of data protection and privacy for Facebook Ireland Limited, dated September 10, 2020. (2) In response toproposed Australian media legislation, Facebook announced that it would stop users in Australia (as well as abroad) from sharing local and international news. Will Easton, *Changes to Sharing and Viewing News on Facebook in Australia*, FACEBOOK (Feb. 17, 2021), https://about.fb.com/news/2021/02/changes-to-sharing-and-viewing-news-on-facebook-in-australia/ [https://perma.cc/7HY7-5AS9]. (3) Ride hailing companies Lyft and Uber threatened to pull out of California if they were forced to classify drivers as employees. Andrew J. Hawkins, *Lyft Joins Uber in Threatening to Pull Out of California Over Driver Status*, VERGE (Aug. 12, 2020, 6:52 PM), https://www.theverge.com/2020/8/12/21365518/lyft-threatens-shut-down-california-q2-2020 [https://perma.cc/8QHV-CXK3]. They eventually defeated the measure via a ballot measure (proposition 22). Kate Conger, *Uber and Lyft Drivers in California Will Remain Contractors*, N.Y. TIMES (Nov. 7, 2020), https://www.nytimes.com/2020/11/04/technology/california-uber-lyft-prop-22.html [https://perma.cc/S5BU-KW5A].

479. GERMAN FEDERAL MINISTRY FOR ECONOMIC AFFAIRS AND ENERGY, PROJECT GAIA-X: A FEDERATED DATA INFRASTRUCTURE AS THE CRADLE OF A VIBRANT EUROPEAN ECOSYSTEM 41 (2019), https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/project-gaia-x.pdf?__blob=publicationFile&v=5 [https://perma.cc/M6HB-KC3J].

providers. Developing countries may benefit if rising competition between different cloud providers not only brings down costs but also increases flexibility about the terms under which this infrastructure is being provided. Similar arguments can be made about reducing dependencies on communication and e-commerce platforms controlled by a small group of U.S.- and China-based technology companies. With proper support and funding, alternative platforms for public communication and e-commerce may emerge.[480]

International development organizations could support initiatives to lessen infrastructural dependencies and related digital inequalities and encourage local digital development and experimentation with public-good-oriented collective governance frameworks. The Universal Postal Union (the "UPU") chose to become a cloud services provider itself in lieu of relying on established commercial cloud services.[481] Working with a local communications provider, the UPU decided to locate the infrastructure and data in Bern, Switzerland (the same country that hosts the UPU Headquarters), under a jurisdiction that fully respects U.N. privileges and immunities.[482] The UPU example is illustrative of the possibility that individual international organizations or consortia could link up to create an interconnected cloud federation.[483] For other types of support for data infrastructures, one might look to the Federated Information System for the Sustainable Development Goals (FIS4SDGs), an initiative led by the Statistics Division of the United

480. France reportedly has plans to launch government versions of Airbnb and Booking.com. Adam Forrest, *France Plans Government Version of Airbnb and Booking.com*, INDEPENDENT (May 21, 2020, 4:37 PM), independent.co.uk/news/world/europe/france-airbnb-government-version-holiday-booking-website-a9526666.html [https://perma.cc/CS89-FSW8]. On public service digital media infrastructure more generally, see Ethan Zuckerman, *The Case for Digital Public Infrastructure,* THE KNIGHT FIRST AMENDMENT INSTITUTE (Jan. 17, 2020), https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure [https://perma.cc/27TW-KVPM]. *See also* Sebastian Benthall & Jake Goldenfein, *Essential Infrastructures*, PHENOMENAL WORLD (July 27, 2020), https://www.phenomenalworld.org/analysis/essential-infrastructures/ [https://perma.cc/3HWK-Z5KA].

481. For a review of cloud computing services in the U.N. system, see generally Jorge T. Flores Callejas & Petru Dumitriu, Joint Inspection Unit, Managing Cloud Computing Services in the United Nations System: Report of the Joint Inspection Unit, U.N. Doc. JIU/REP/2019/5 (2019).

482. *Id.*

483. This might enable IOs, as cloud providers, to link up to GAIA-X or to other regional data sharing infrastructures. The European Commission's European Strategy for Data contemplated Memoranda of Understanding with EU Member States, starting with those having existing cloud federation and data-sharing initiatives. *See A European Strategy for Data*, at 18, COM(2020) 66 final.

Nations Department of Economic and Social Affairs (DESA) in partnership with Esri, a company that supplies geographic information system software and geodatabase management applications.[484] The FIS4SDGs initiative is based on the principle of "national ownership," with the National Statistical System implementing "internationally agreed standards" for the production and dissemination of data and statistics.[485] In this context, the National Statistical Offices are envisioned to have a leadership role, "[c]oordinating the [national statistical systems] and improving cooperation between data producers[, s]upporting statistical work of line ministries and other entities[, and] validating data from different sources for consistency, accuracy and reliability[.]"[486] The Country Data Hubs will contain geospatially enabled datasets pertaining to specific Sustainable Development Goals (SDG) indicators as well as interactive analytical visualization and communication applications, such as Story Maps.[487] Through the federated architecture, Country Data Hubs can share SDG-relevant data with each other, "enabling users to not only access the data they need when they need it, but also ensure the traceability and accountability of the data, which is maintained at its source."[488] Data hubs of international agencies will aggregate data from national data hubs and allow users to access harmonized data.[489] The United Nations has already introduced the global Open SDG Data Hub,[490] and a number of countries have also launched their own SDG data hubs.[491] This type of data infrastructure, if executed well, can provide a balance

484. U.N. DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, FEDERATED INFORMATION SYSTEM FOR THE SDGS: A PLATFORM FOR THE SHARING OF NATIONAL AND GLOBAL STATISTICAL AND GEOSPATIAL DATA FOR THE 2030 AGENDA, https://www.unescap.org/sites/default/files/Session_4_Intro_to_federated_information_system_for_the_SDGs_WS_National_SDG_10-13Sep2019.pdf [https://perma.cc/ER8K-9MLZ].

485. Presumably, these standards would be agreed upon under the auspices of the U.N. Statistical Commission.

486. U.N. DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, *supra* note 484, at 12. It is emphasized that the data published by data hubs will be "authoritative data." *Id*. at 18.

487. FEDERATED INFORMATION SYSTEM, *supra* note 93.

488. *Id*.

489. *Interconnected Data Hubs and Public Participation: The Data Revolution Is Underway*, U.N. DEP'T OF ECON. & SOC. AFF. (Mar. 9, 2018), https://www.un.org/development/desa/en/news/statistics/data-revolution-underway.html [https://perma.cc/PF26-CGRH].

490. *Welcome to the Open SDG Data Hub*, UNITED NATIONS STATISTICS DIVISION, http://www.sdg.org/ [https://perma.cc/76FU-H9RE].

491. For a short case study on the UAE's SDG Data Hub, see *UAE Data Hub Drives Sustainability Goals*, ESRI, https://www.esri.com/en-us/industries/government/departments/lighthouse-case-study [https://perma.cc/BB8L-E5GZ].

between local control over data production processes and collective and deliberative decisions over data use on the one hand, and transnational data sharing for collectively agreed-upon purposes on the other.

Private commercial actors are not necessarily excluded from the data infrastructures created and supported by either local constituencies or international organizations. It is imperative, however, that their engagement is attuned to the dynamics of infrastructural control over data, as we have highlighted throughout this Article. The contractually agreed terms ought to be mindful of dependencies (ensuring interoperability to enable switching as needed) and about questions of data generation and control. Negotiating such terms often necessitates confronting a reality of severe power imbalances. Fostering public data infrastructures may not only produce competitive pressures on commercial actors, but it may also set standards for best practices and terms under which data is generated and used. Furthermore, the collective pooling of resources might rebalance the negotiating power of public entities. In this respect, it is worth noting that international organizations have vast amounts of diverse data that might be usefully pooled together and, with appropriate safeguards, be deployed to create global public-private data sharing platforms that would enable differentiated access to data to different actors.[492] The collective political power of international organizations might also be leveraged to set fair terms for participation of commercial actors.

Lastly, like international institutions, international law itself could also be mobilized in the quest to reclaim infrastructural control over data. Estonia is often hailed as a role model for successful digital transformation, as the government decided to introduce novel digital infrastructures for identification and governmental services, as well as encouraged widespread adoption of digital technologies by citizens and businesses. When faced with the choice of where to store governmental data to shield it from cyberattacks, Estonia initially considered partnering with a U.S.-based commercial cloud provider. Eventually, however, it settled on a "data embassy," a data center located physically in Luxembourg and protected by an agreement between the two governments, with the necessary technology provided by various private sector entities.[493] Estonia's journey towards its

492. For discussion of differentiated data access, see *infra* Section III.D.

493. Yuliya Talmazan, *Data Security Meets Diplomacy: Why Estonia Is Storing Its Data in Luxembourg*, NBC NEWS (June 25, 2019, 12:33 PM), https://www.nbcnews.com/news/world/data-security-meets-diplomacy-why-estonia-storing-its-data-luxembourg-n1018171 [https://perma.cc/ZNY2-3JPD].

"data embassy" is instructive. It highlights the salience of the bond between infrastructures and attendant legal structures under public international law (government-to-government) and transnational private law (between the government and businesses) and represents an instantiation of creative thinking in light of the country's particular geopolitical, economic, and regulatory context.

C. Demanding Transparency

One of the key preconditions for regulating infrastructural control over data and for planning digital policy more generally is knowledge about how relevant actors exercise such control, what data they generate and accumulate and through what means, and how they use infrastructures to entrench their market positions and cement control over data infrastructures and data. These questions are related to, but also different from, the dominant discourse about the opacity and inscrutability of algorithms. That discourse typically is concerned with the use of artificial intelligence/machine learning algorithms for automated and/or predictive decision-making, where the terms and assumptions baked-into the algorithms are not easily understood by humans who are ultimately impacted by those decisions.[494] Antitrust and competition law investigations, such as those carried out in the United States, the European Union, and other jurisdictions, may reveal the extent of control over data and related data-generating and -transacting processes to authorities, albeit only to the extent to which such disclosures are necessary for such investigations.[495] This type of disclosure is different from (and falls short of) demanding transparency about corporate data generation to reveal infrastructural power and resulting data control and datafication power asymmetries for the purposes of public oversight, contestation, and deliberation. Transparency over data inequality entails asking how much and what kind of data corporations control as well as demanding the accompanying metadata outlining the context within which the data was collected.[496] Such demands will not only facilitate ascertaining the economic value of the accumulated data, but also will provide insight into the process of datafication itself.

---

494. *See generally* PASQUALE, *supra* note 121; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018).

495. On the potential and limitations of such inquiries and their remedies, see *supra* Section II.D.

496. *See* Timnit Gebru et al., *Datasheets for Datasets*, arXiv (2020), https://arxiv.org/abs/1803.09010 [https://perma.cc/DXB4-CHLC].

Despite the widespread assumption that platform companies control vast amounts of data, surprisingly little is known about *how much* data they actually control, how the data they control is being generated, and what economic value "their" data holds. The inability to account for data as an economic asset is, at least in part, a function of contemporary accounting standards, which do not account for much data, despite the widespread belief that data is becoming companies' most important asset.[497] Jonathan Haskel and Stian Westlake have described how the "knowledge economy" increasingly relies on investment in research and development, which leads to ideas that may or may not be protected under intellectual property law and are only imperfectly accounted for under existing accounting standards.[498] The increasing salience of data for successful businesses compounds the problem, as the creation of data is not necessarily commensurate with the investment undertaken to create the data. No one knows how much data the world's leading technology companies control. As long as this is the case, scholars and policymakers alike are left with theoretical arguments and mere guesswork about the extent of contemporary data control asymmetries and the degree to which these stem, as we posit, from control over data infrastructures.

The "accounting for data" problem extends to conventional economic and trade statistics which largely do not account for data.[499] Instead, these statistics tend to measure the excesses of the digital economy, where control over data gets commercialized, especially through advertising.[500] This is not only a problem for those who are tasked with advising governments on the state of the domestic and global economy, or those who rely on this information for their financial and commercial strategies; it also severely undermines the political discourse around digital development and which strategies to pursue. Critics of contemporary digital development strategies struggle to substantiate their arguments in the absence of reliable data. Proponents of digital development that adhere to the Silicon Valley Consensus likewise struggle to make their case for unrestricted data flows and against data localization measures. They often rely on questionable proxies (e.g., bandwidth expansion) or general

497. *See* discussion *supra* Section I.A.

498. *See generally* JONATHAN HASKEL & STIAN WESTLAKE, CAPITALISM WITHOUT CAPITAL: THE RISE OF THE INTANGIBLE ECONOMY (2017).

499. *See generally* Dan Ciuriak, Unpacking the Valuation of Data in the Data-Driven Economy (NYU Conference on Global Data Law, 2019), https://ssrn.com/abstract=3379133 [https://perma.cc/AK3X-6JWD].

500. *See generally* MARIANA MAZZUCATO, THE VALUE OF EVERYTHING: MAKING AND TAKING IN THE GLOBAL ECONOMY (2018).

assumptions (e.g., about regulatory uncertainty) in the absence of more precise and differentiated data about who has what kind of data (data control) and between whom data flows.[501]

Some platform companies have begun to make certain kinds of data available to researchers.[502] Such efforts at "voluntary" data sharing, sometimes branded as "data philanthropy,"[503] are inherently one-sided, as data demanders often do not even know which data exists and data holders often hide behind data protection laws to claim that sharing is impossible (without detailing why).[504] In addition, corporate actors deploy a variety of legal tools—including trade secrets protection, non-disclosure agreements, property claims, and invocations of privacy concerns—to prevent disclosures.[505] For

501. The European Union launched a project on data flow monitoring to map current data stocks and flows within EU territory. The effort is, however, so far entirely based on voluntary surveys and hence unlikely to produce an accurate picture of data flow reality. *See The European Data Flow Monitoring*, EUR. COMM'N (Mar. 9, 2021), https://digital-strategy.ec.europa.eu/en/policies/european-data-flow-monitoring [https://perma.cc/NC59-7L62].

502. See, for example, Google Research's release of the Objectron Dataset, a machine-learning dataset for 3D object recognition. Adel Ahmadyan and Liangkai Zhang, *Announcing the Objectron Dataset*, GOOGLE AI BLOG (Nov. 9, 2020), https://ai.googleblog.com/2020/11/announcing-objectron-dataset.html [https://perma.cc/4STP-SMNC]. Alibaba released datasets about the servers and running tasks in its production clusters. Alibaba Tech, *Open Season for Research: Alibaba Releases Cluster Data from 4000 Servers*, HACKERNOON (Jan. 22, 2019), https://hackernoon.com/open-season-for-research-alibaba-releases-cluster-data-from-4000-servers-12d013bd6b4e?gi=5836a4e8dafa [https://perma.cc/7KKT-V88N]. In 2021, Twitter announced it will offer a full history of its full-archive search endpoint to any researcher or developer who applies as part of the launch of a new academic research track. Nick Statt, *Twitter Is Opening Up Its Full Tweet Archive to Academic Researchers for Free*, THE VERGE (Jan. 26, 2021, 2:00 PM), https://www.theverge.com/2021/1/26/22250203/twitter-academic-research-public-tweet-archive-free-access [https://perma.cc/2VFX-47AH].

503. *See generally* Yafit Lev Aretz, *Data Philanthropy*, 70 HASTINGS L.J. 1491 (2018–2019).

504. *See generally* Mathias Vermeulen, *The Keys to the Kingdom. Overcoming GDPR-Concerns to Unlock Access to Platform Data for Independent Researchers*, OSF PREPRINTS (Nov. 27, 2020).

505. For example, in August 2021, Meta (formerly Facebook) disabled the accounts, apps, pages and platform access associated with NYU's Ad Observatory Project, which sought to understand the role of political advertising on Facebook by creating a browser extension (AdObserver) to allow Facebook users to volunteer "their" data in anonymous fashion. NYU AD OBSERVATORY PROJECT, https://adobservatory.org [https://perma.cc/UZQ5-8EG4]. Meta justified its actions on the grounds that researchers' actions jeopardized people's privacy. Mike Clark, *Research Cannot Be the Justification for Compromising People's Privacy*, META (Aug. 3, 2021) https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy [https://perma.cc/NYN8-EKE7]. The U.S. Federal Trade

related reasons, we also caution against voluntary "data reports." Companies have issued "transparency reports" about certain data-related activities, which are often transparent in name only, as they hide important information in "aggregated data," are short on explanations, and reflect companies' choices about what to report.[506]

Certain data protection laws give individuals the (often costly and resource-intensive) right to inquire about the personal data that companies hold about them.[507] Data protection and cybersecurity laws require the disclosure of certain cyber incidents and data breaches.[508] But there is currently no law that systematically addresses the lack of transparency when it comes to control over data in the digital economy. Regulatory bodies thus far appear to have been reluctant to demand such transparency. Statistical units of governments and international organizations have begun to realize that private actors might have superior data, but, to our knowledge, they have not thus far demanded disclosure and sharing of such data.

We suggest that more forceful governmental intervention may be needed to remedy data inequality. The transparency requirements that we envision would depart from the individual rights-based or incident-based approach under which individuals can demand access to "their" data or companies' need to disclose information about a data breach or cybersecurity incident. Instead, controllers of data infrastructures above a certain threshold (e.g., determined, by market

---

Commission issued a strongly worded rebuke, accusing Facebook of invoking privacy commitments as pretext to shield its political advertising business from public scrutiny. Letter from Acting Director of the Bureau of Consumer Protection Samuel Levine to Facebook (Aug. 5, 2021), https://www.ftc.gov/news-events/blogs/consumer-blog/2021/08/letter-acting-director-bureau-consumer-protection-samuel [https://perma.cc/8CJT-QXVG]. More generally, see discussion *supra* Sections II.A–C about how these legal technologies may contribute to data inequality.

506. On the virtues of enhanced transparency in the context of platform governance, see generally Tarleton Gillespie, *Regulation of and by Platforms*, *supra* note 98, ch. 14; Daphne Keller & Paddy Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation*, *in* SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM (N. Persily & J. Tucker eds., forthcoming), https://ssrn.com/abstract=3504930 [https://perma.cc/8LAA-GUNH] (surveying different types of data disclosures accessible to researchers). *But see* Monika Zalnieriute, *"Transparency-Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism*, 8 CRITICAL ANALYSIS OF L. 39 (2021) (critiquing "transparency-washing" as a strategy of obfuscation and redirection away from more substantive and fundamental questions).

507. *See* Jef Ausloos & Michael Veale, *Researching with Data Rights*, TECH. & REGUL. 136, 156 (2021) (showing how researchers can leverage data rights to gain access to enclosed datasets).

508. Mark Verstraete & Tal Zarsky, *Optimizing Breach Notifications*, 2021 UNIV. OF ILL. L. REV. 803, 809–30 (2021).

capitalization, market share, number of users, or type of data) would be required to disclose how much data of what kind they control and through what means (i.e., which data infrastructures).[509] Such a requirement would be akin to the financial disclosure requirements imposed on financial institutions of systemic stature.[510] While this type of intervention would certainly impose considerable compliance costs, and may even require investment in infrastructures necessary to make such determinations (including to protect individuals' privacy as far as personal data is concerned), we do not believe that imposing such requirements on the largest providers of data infrastructures and controllers of data would render their operations unprofitable. Only the world's most powerful regulators with commensurate market power (and possibly multilateral standard-setting bodies) will be able to demand and effectively enforce this level of commitment towards transparency.[511] Ideally, such measures would create positive spill-over effects for others if companies are forced (or decide) to implement heightened transparency requirements globally.[512] Increased transparency could also create opportunities for activists to challenge the exploitative datafication on a population-level scale.[513] In the

509. For a similar proposal based on GDPR access to data rights, see generally RENÉ L. P. MAHIEU & JEF AUSLOOS, LAWARXIV, RECOGNISING AND ENABLING THE COLLECTIVE DIMENSION OF THE GDPR AND THE RIGHT OF ACCESS (2020). As with other data protection-based approaches, their proposal is limited to personal data. *See* discussion *supra* Section II.C on this limitation.

510. For similar ideas drawing on financial market regulation, see generally Salome Viljoen & Sebastian Benthall, *Data Market Discipline: From Financial Regulation to Data Governance*, 8 J. INT'L & COMP. L. 459 (2021).

511. See, for example, the U.S. Consumer Financial Protection Bureau (CFPB) requesting information on the business practices of large technology companies operating payments systems in the United States. Press Release, CFPB Orders Tech Giants to Turn Over Information on their Payment System Plans (Oct. 21, 2021), https://www.consumerfinance.gov/about-us/newsroom/cfpb-orders-tech-giants-to-turn-over-information-on-their-payment-system-plans [https://perma.cc/GQ3R-N3LY]; *see also* Caroline A. Crenshaw, *Mind the (Data) Gaps*, KEYNOTE ADDRESS AT THE 8TH ANNUAL CONFERENCE ON FINANCIAL MARKET REGULATION (CFMR) (May 14, 2021), https://www.sec.gov/news/speech/mind-the-data-gaps [https://perma.cc/UCJ7-8Q57] (calling for more expansive data disclosures to understand markets and associated risks).

512. This is the dynamic that Anu Bradford has theorized as the "Brussels Effect." BRADFORD, *supra* note 298, ch. 2.

513. For an example of an effective challenge to the exercise of data-based administrative power by a state, see NJCM v. the Netherlands (SyRI), ECLI:NL:RBDHA:2020:1878. *See also* Christiaan van Veen, *Landmark Judgment from the Netherlands on Digital Welfare States and Human Rights*, OPENGLOBALRIGHTS (Mar. 19, 2020), https://www.openglobalrights.org/landmark-judgment-from-netherlands-on-digital-welfare-states [https://perma.cc/9VWN-MFD3].

meantime, and concordantly, other actors can take steps in the same direction by demanding transparency whenever they negotiate contracts with data infrastructure providers, and by banding together, if necessary, to increase their collective bargaining power.

The proposed DMA and DSA may mark a turning point, as they impose significant transparency obligations, mandating data access for supervisory authorities.[514] The DSA also includes data access provisions for vetted researchers.[515] As the DSA proposal notes, "Investigations by researchers on the evolution and severity of online systemic risks are particularly important for bridging information asymmetries and establishing a resilient system of risk mitigation . . . "[516] Unlike much of the extant law discussed in Part II, these recent EU initiatives focus not just on data as an object but also on data infrastructures themselves. For example, the proposed DMA requires transparency about certain processes of datafication.[517] Yet, these initiatives also illustrate the limits of transparency in regulating infrastructural control, given that data infrastructures are notoriously complex and contextual.[518] Understanding how the power to datafy is exercised (or how a platform can be deployed to generate more data)

514. *See supra* notes 375–376 and accompanying text. The DSA proposal, *supra* note 369, recital 99, states:

> [T]he Commission should have access to any relevant documents, data and information necessary to open and conduct investigations and to monitor the compliance with the obligations laid down in this Regulation, irrespective of who possesses the documents, data or information in question, and regardless of their form or format, their storage medium, or the precise place where they are stored.

515. New transparency obligations are also under consideration in the U.S. Congress. *See* Tara Wright, *The Platform Transparency and Accountability Act: New Legislation Addresses Platform Data Secrecy*, STAN. CYBER POL'Y CTR. (Dec. 9, 2021), https://cyber.fsi.stanford.edu/news/platform-transparency-and-accountability-act-new-legislation-addresses-platform-data-secrecy [https://perma.cc/E7AJ-VU2Y].

516. DSA Proposal, *supra* note 369, recital 64. The DSA requires very large online platform companies to provide the supervising authority access to data that is "necessary to assess the risks and possible harms brought about by the platform's systems, data on the accuracy, functioning and testing of algorithmic systems for content moderation, recommender systems or advertising systems, or data on processes and outputs of content moderation or of internal complaint-handling systems . . . " *Id.*

517. For example, the proposed DMA says that gatekeepers should:

> [*A*]*t least* provide a description of the basis upon which profiling is performed, including whether personal data and data derived from user activity is relied on, the processing applied, the purpose for which the profile is prepared and eventually used, the impact of such profiling on the gatekeeper's services, and the steps taken to enable end users to be aware of the relevant use of such profiling, as well as to seek their consent.

DMA, *supra* note 370, recital 61 (emphasis added).

518. *See* discussion *supra* Section I.C.

requires access not only to the existing data, but also to the processes—technical, organizational, and social—through which decisions about data generation are made. Regulators could consider deploying ethnographers as part of the auditing processes to gain insights into these processes.[519]

D. Pooling and Differentiating Access to Data

In the preceding Sections, we focused mainly on state-level regulatory interventions and the role of intergovernmental organizations. In this Section, we consider different local practices and governance arrangements and their potential to foster sustainable digital development from the bottom-up and to reallocate the power to decide, in a more participatory fashion, what data is generated, for what purpose, and on what terms. We consider the development of data infrastructures more attuned to local contexts and governed by local communities. The nascent models we consider in this Section underscore that locality is not delineated by geographical location.[520] Instead, the locality represented here is relational and relative. In some instances, "local" practices are positioned in relation to national or global contexts; in others, they emphasize a particular community and are "localized" in terms of common interests or goals. Local ownership, sourcing, or practices are not an end in themselves, as local practices can be exclusionary or even oppressive as well.

We draw inspiration from different emerging models of collective governance over data and data infrastructures, often billed as "data cooperatives," "data collectives," "data commons" or "data trusts".[521] The emergence of such initiatives has been sporadic, and their success is difficult to assess for a variety of reasons: in part because of their hyper-local nature, their novelty, and the difficulty of

519. Corporations frequently commission ethnographers to understand how their products and services are used. Leslie Brockow, *Ethnography in Action at Wells Fargo*, MIT SLOAN MGMT. REV. (Mar. 30, 2014), https://sloanreview.mit.edu/article/ethnography-in-action-at-wells-fargo [https://perma.cc/AH9C-6487]; Michael Fitzgerald, *Corporate Ethnography*, MIT TECH. REV. (Nov. 17, 2005), https://www.technologyreview.com/2005/11/17/230047/corporate-ethnography [https://perma.cc/TE78-EJVV].

520. On the different meanings of "local" in the context of knowledge and data, see LOUKISSAS, *supra* note 438.

521. *See* Ada Lovelace Institute & UK AI Council, EXPLORING LEGAL MECHANISMS FOR DATA STEWARDSHIP (Mar. 2021), https://www.adalovelaceinstitute.org/project/legal-mechanisms-for-data-stewardship-working-group [https://perma.cc/59WU-Q8QK]; Bianca Wiley & Sean McDonald, *What Is a Data Trust?*, CIGI ONLINE (Oct. 9, 2018), https://www.cigionline.org/articles/what-data-trust [https://perma.cc/6MUF-Z3MQ].

predicting their long-term sustainability and outcomes. The legal environment in which these initiatives operate may be changing as well. For example, the proposed European Data Governance Act tries to create more favorable conditions for data cooperatives and other data sharing intermediaries.[522]

Although emerging initiatives for collective data governance often share similar labels, they vary in form, scope, and governance type.[523] Some aim to incentivize the pooling of data for public purposes, while giving data contributors various degrees of control and choices about how their data is used. For example, Salus, a non-profit Barcelona-based citizen data cooperative for health research, designed a license that allows data to be donated for research purposes under set conditions.[524] Similarly, the Driver's Seat Cooperative aims to give ride-sharing drivers an opportunity to monetize their driving data through sales of insights to city agencies, so they can make better transportation planning decisions.[525] Proceeds from sales are shared among the driver-owners via dividends.[526]

Another set of cooperatives and other community-based arrangements have arisen specifically to counter the infrastructural control of large commercial cloud providers.[527] For some initiatives,

---

522. Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act) COM/2020/767 final (2020); *see also* Sean McDonald, *A Novel, European Act of Data Governance*, CIGI ONLINE (Dec. 15, 2020), https://www.cigionline.org/articles/novel-european-act-data-governance [https://perma.cc/DH2K-J6FK].

523. On the idea of data cooperatives, see generally Alex Pentland et al., *Part I: The Human Perspective: New Types of Engagement*, *in* BUILDING THE NEW ECONOMY: DATA AS CAPITAL (Alex Pentland, Alexander Lipton & Thomas Hardjono eds., 2020).

524. Data must be for use in health research to be used by non-profit institutions that openly share the results of their research while anonymizing the data at the highest possible level. Use is allowed until data donors withdraw their permission. SALUS COOP, https://www.saluscoop.org/manifiesto [https://perma.cc/7LWL-ZMBN]. In 2020, Salus created the Cooperative COVID Cohort project (CO3) to create a cohort of citizen data donors for research on COVID-19. *Salus CO3: Cooperative COVID Cohort*, SALUS COOP, https://www.saluscoop.org/proyectos/co3 [https://perma.cc/Y6EJ-7FY8].

525. See the self-description at DRIVER'S SEAT, https://driversseat.co [https://perma.cc/4F5B-QA5M].

526. *Id.*

527. For example, CoBox offers a distributed, encrypted, offline-enabled data hosting cloud platform. Its stated aim is "to facilitate a transition away from giant data centers, huge storm clouds, towards a vision of cloud infrastructure that is light, distributed, and importantly, is offline-first . . . CoBox is the beginning of a sovereign commons-based data infrastructure and a co-operative distributed cloud architecture." COBOX, https://cobox.cloud/how-it-works [https://perma.cc/MZ75-M6P8].

collective governance is a core feature. For example, CommonsCloud is a cloud service set up specifically to function as an alternative to Google Drive, Amazon Drive, Microsoft One, Apple iCloud, and Dropbox.[528] It is a self-managed community whose users can also become consumer partners of another cooperative (femProcomuns) and participate in the governance of both that cooperative and CommonsCloud.[529] Similarly, Framasoft, a nonprofit network of projects headquartered in France, hosts a project to "de-google-ify" the internet by offering free alternative services. It has an elaborate governance framework that resembles those used in open-source software communities.[530]

One of the first and the best known examples of collective data-sharing arrangements that combine data pooling with collective governance is Project DECODE, which is connected to broader efforts to reimagine democratic governance in digitally-mediated cities.[531] Project DECODE was a pilot that aimed to give individuals in different communities control over how and on what terms data that is generated and gathered by apps, interconnected devices, and sensor networks in cities, can be made available for broader communal use, with appropriate privacy protections.[532] It was rolled out in the cities of Barcelona and Amsterdam. In Barcelona, residents used DECODE

528. Commonscloud.coop, project of a Catalan community, is specifically billed as an alternative to corporate storage clouds such as Google Drive and Dropbox. CommonsCloud provides free software, as well as a platform that facilitates community conversation around services provided by the project. COMMONSCLOUD.COOP, https://www.commonscloud.coop/projecte [https://perma.cc/Q2T8-C4M6].

529. FemProcomuns is a "non-profit and social initiative, worker and consumer multi-stakeholder cooperative, created in Catalonia in 2017, with the aim of consolidating a commons ecosystem, based on the principles of open cooperativism, community self-management, human, ecological, and economic sustainability, shared knowledge and replicability." It provides organizational, operational and governance functions to cooperatives. FEMPROCOMUNS, https://femprocomuns.coop/about-femprocomuns/?lang=en [https://perma.cc/YUP2-PASL].

530. A list of the offered services is available here: *De-google-ify Internet*, FRAMASOFT, https://degooglisons-internet.org/en/list [https://perma.cc/GR3Y-4FRE]. The governance structure is explained here: *Association*, FRAMASOFT, https://framasoft.org/en/association/ [https://perma.cc/5B9Y-PX3Y].

531. DECODE, *supra* note 92; Evgeny Morozov & Francesca Bria, *Rethinking the Smart City: Democratizing Urban Technology*, ROSA LUXEMBURG STIFTUNG (2018); Bianca Wiley, *Searching for the Smart City's Democratic Future*, CTR. FOR INT'L GOVERNANCE INNOVATION (Aug. 13, 2018), https://www.cigionline.org/articles/searching-smart-citys-democratic-future [https://perma.cc/YJ8C-CH28].

532. Citizens can set the anonymity level via the DECODE app so that they cannot be identified without explicit consent. In this way, they can keep control over data once they share it for the communal purposes. *See supra* note 92 and accompanying text.

technology (a combination of blockchain and attribute-based cryptography) to share encrypted data anonymously within their community in response to community concerns that data from environmental sensors placed throughout the neighborhood (which recorded noise levels, pollution, temperature, humidity, etc.) might reveal sensitive information or be misused otherwise. DECODE also provided training and support to the individuals on how data could be gathered, analyzed, and used to improve city services.

The data shared by citizens in the DECODE pilots was meant to "[integrate] with the Barcelona City Hall digital infrastructures: the data lake CityOS, the IoT open sensor network Sentilo, Barcelona open data portal and the digital democracy platform Decidim."[533] Such integration was possible only after the Barcelona City Council released a new Digital City Plan, with an ethical data strategy.[534] Barcelona also revised procurement deals between the City Hall and its private sector providers and included "data sovereignty" clauses in public procurement contracts, requiring suppliers working for the city to provide the data they gather to deliver services in machine-readable format, thus enabling the release of such data as open data and allowing communities to benefit from it as well.[535]

Whether initiatives inspired by Project DECODE succeed in generating enough uptake and participation to achieve their aims remains to be seen.[536] Nonetheless, these examples offer an alternative pathway to developing data infrastructures. One can imagine, for example, a series of data collectives linked up along sectoral lines (e.g., to expand the sharing of health or transportation data beyond an immediate community), common values (e.g., the promotion of open, free, and decentralized infrastructure), geographic proximities (e.g., cities in Europe), and/or along other dimensions. Communities could band together with other similarly predisposed publics to develop the necessary technologies, to exchange experiences and ideas, and to

533. *Common Knowledge: Citizen-Led Data Governance for Better Cities*, DECODE (Jan. 2020), https://decodeproject.eu/publications/common-knowledge-citizen-led-data-governance-better-cities.html [https://perma.cc/C7SJ-RAPZ].

534. BARCELONA DIGITAL CITY: PUTTING TECHNOLOGY AT THE SERVICE OF PEOPLE 4 (2015), https://ajuntament.barcelona.cat/digital/sites/default/files/pla_barcelona_digital_city_in.pdf [https://perma.cc/QAB2-8WR6].

535. For a discussion of distributional effects of open data, see discussion *supra* Section II.B.

536. Public support of the kind illustrated by the Barcelona case is critical to the uptake and sustainability of local infrastructures. Without it, many worthwhile initiatives will likely be abandoned due to limited acceptance. *See, e.g.*, THEGOODDATA, https://www.thegooddata.org [https://perma.cc/YE3G-JZ2V].

develop commensurate public governance mechanisms attuned to their specific conditions.[537]

Many of the collective initiatives aim to create new data infrastructures, not only to enable the pooling of and control over data generation, but also to release pooled data to the public as "open data".[538] Making certain datasets publicly available can have substantial public benefit. For example, having access to different datasets on COVID-19 infections, hospitalizations, and deaths enabled different platforms to compile real-time COVID-19 profiles for countries (e.g., the dashboard created by Johns Hopkins University[539]) and unveiled opportunities for scientific examination and collaboration.[540] At the same time, open data, at least as practiced so far, may not be an effective tool to address data inequality.[541] Critically, thus far, calls to "open" data have been mainly directed at governments.[542] Calls to "open" up privately-held data have been comparatively rare. Demands for open public data—including those in "digital trade" agreements—are often propelled by narratives according to which they provide opportunities for small and medium businesses.[543] Yet, the extent to which access to open governmental data increases the opportunity of smaller enterprises to compete with corporations that control expansive data infrastructures is unclear. Those who control the means of data generation on large scales gain access to the same "open" data, once made available by governments, as everyone else. Given that one value proposition of datafication lies in producing insights from data, which often requires aggregation with other datasets, open data tends to privilege tech-savvy users, who have both the capacity and resources to integrate and analyze different datasets, as well as the access to other datasets (open and closed),

537. Indeed, one could envisage data being the catalyst for new types of transnational movements that coalesce around common interests (e.g., labor exploitation).

538. Although there are multiple meanings of "open," as used in this context, "open data" refers to data made publicly available under an open data license or data that is released into public domain. This is legally complicated because data—unlike software—does not acquire copyright protection as easily. *See supra* Section II.B.

539. *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*, JOHNS HOPKINS UNIV., https://coronavirus.jhu.edu/map.html [https://perma.cc/C6LP-86HS].

540. Junaid Shuja et al., *COVID-19 Open Source Data Sets: A Comprehensive Survey*, APPLIED INTEL., 1–8 (Sept. 2020).

541. *See* discussion *supra* Section II.B.

542. *See generally* GRAY, *supra* note 87.

543. *See, e.g.*, Joel Gurin, *Driving Innovation with Open Data*, THE FUTURE OF DATA DRIVEN INNOVATION 55 (2014).

which enhance the re-usability value of open data to them. Thus, perhaps counterintuitively, open data can further empower those who already have enhanced access to data, including large data companies like Google and Alibaba, thus failing to correct the asymmetry in data distribution.

Open data also has the potential to exacerbate the inequality of the power to datafy. Decisions about which data is made open and under what conditions are usually made by the data holders and the processes according to which these determinations are being made (and with what considerations and motivations in mind) are often not disclosed. By the time data is opened up, decisions about how the phenomenon it purports to represent was defined, measured, collected, and processed have already been made, as have choices about what features of the phenomenon *not* to datafy or represent. These choices—and all the biases they contain—are reproduced and entrenched as data becomes released for public use.

These effects could, of course, be mitigated through regulatory design and governance arrangements—for example, by fostering participatory multi-stakeholder processes for data collection. The Open Government Partnership purports to take this approach.[544] If open data is not confined to output data and metadata—instead encompassing a full disclosure of methods and sources through which data was selected and processed—crucial context comes into focus, which may then enable productive contestation.[545] Similarly, data can be made more useable and legible to different audiences through choices of format in which it is made available (e.g., machine-readable, narrative-style or other formats for low-tech engagement) and by simplifying the means through which it can be accessed (e.g., download from a website, access through APIs etc.), thereby reducing the need to rely on platforms as gateways to open data. Open data is often first consumed and subsequently made legible by intermediaries, which might include librarians, journalists, and nonprofits.[546] However, efforts (and funds) directed at making data open are rarely also directed at supporting and sustaining the intermediaries who lend

544. OPEN GOVERNMENT PARTNERSHIP, https://www.opengovpartnership.org/process/ [https://perma.cc/C6YH-DZAP].

545. For an example of this approach, see the WomanStats Project: THE WOMANSTATS PROJECT, https://www.womanstats.org/aboutoverview.html [https://perma.cc/SD45-NYGE].

546. On the role of infomediaries, see generally Ricardo Ramírez, Balaji Parthasarathy & Andrew Gordon, From Infomediaries to Infomediation at Public Access Venues: Lessons from a 3-Country Study 124 (Proceedings of the Sixth International Conference on Information and Communication Technologies and Development, 2013).

vital support to enable a wider use of open data by less-resourced constituencies.

Still, even with these mitigation strategies, from the perspective of data as an economic resource, there is little evidence that unconditional opening of public data in and of itself can address unequal control over data or redistribute opportunities for monetization. We highlight these challenges not to disparage the value of open data, but instead to encourage a more nuanced and differentiated consideration of how, for what purpose, and under what conditions access to data is provided. If one seeks to confront this form of data inequality, one cannot assume that interests of all stakeholders are necessarily aligned (though they sometimes are) and that all stakeholders ought to be treated in the same way. For this reason, developing differentiated and conditional "data sharing" infrastructures attuned to data inequality might be a worthwhile pursuit for smaller markets and economies. In this context, we suggest building on Lisa Austin's and Davie Lie's idea of "Safe Sharing Sites" as infrastructural and legal assemblages that make data more regulatable.[547] While their focus is on facilitating data-sharing, while protecting the data privacy and security of individual data subjects, their approach can be expanded to design data-sharing infrastructures, including for non-personal or aggregated data, that cater to additional and different concerns, including those around data inequality. This solution is not ready-made or straight-forward. The key point, however, is that addressing data inequality through "data sharing" requires careful consideration of infrastructural and legal elements, and that the "safe sharing site" idea encourages thinking along these dimensions. In this regard, we depart from much of the "open data" discourse, which tends to be focused on the licensing terms under which data is being made available, as well as "data collaborative" ideas, which are often centered on contractual solutions.[548] While these interventions are valuable in themselves, they are unlikely to address data inequality dimensions because they tend to disregard the severe gradients of disproportionate power that stem from asymmetric control over data and data infrastructures, as discussed throughout this Article. Those who already control data and data infrastructures stand most to gain from "open data" initiatives and can also dictate the terms in "data collaboratives."[549]

---

547. *See generally* Lisa Austin & David Lie, *Safe Sharing Sites*, 94 N.Y.U. L. REV. 4 (2019)

548. See, for example, the Contracts for Data Collaboration (C4DC) project, www.contractsfordatacollaboration.org [https://perma.cc/ZZ3Z-2MVM].

549. *See generally* GRAY, *supra* note 87.

In terms of differentiation, designers of data-sharing infrastructures can choose and decide who gets access to data and on what terms. This design power needs to be checked through appropriate governance mechanisms, but it can also be mobilized for public benefit: those who commit to using data for non-commercial purposes, for example, could be granted access to data for free, while those with commercial use-cases could pay a fee, which could be dependent on business size or commensurate with eventual profit. Logically, future data-sharing outside the "safe sharing site" would need to be constrained and policed to retain regulatory control over the practice. Importantly, the safe sharing site would not make data available wholesale, but rather circumscribed to the relevant use on a case-by-case basis. Comparable solutions are already in place for differential access to data in various contexts. One example is machine learning when the quality of the models can be trained and tested without gaining access to the actual testing data.[550] Ideas to share (only) the insights from machine learning (rather than the underlying data) point into a similar direction.[551]

Access to the public data sharing infrastructure could be conditioned on a range of regulatory demands: for example, requiring adherence to data protection and privacy policies, commitment to respect, and fostering spaces for deliberation and contestation of proposed data uses; demanding payment of fees to fund the infrastructure; or even requiring tax residency.[552] Naturally, such conditionalities will only be effective if the expected benefit that can be derived from accessing the data outweighs the costs that these conditionalities impose.[553]

Importantly, public data sharing infrastructures discussed in this Section can involve and benefit a range of public and private actors, but control over the platforms' legal and infrastructural design is crucial. For this reason, it is of paramount importance to develop

550. For an explanation of the practice of "querying data," see, for example, *What Is Differential Privacy in Machine Learning*, MICROSOFT (Nov. 5, 2021), https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy [https://perma.cc/PPJ7-TELM].

551. For such ideas, see Michal Gal & Nicolas Petit, *Radical Restorative Remedies for Digital Markets*, 37 BERKELEY TECH. L.J. (forthcoming 2022).

552. Denmark blocked firms registered in tax-havens from receiving state aid during the COVID pandemic. *See* Nikolaj Skydsgaard, *Denmark Blocks Firms Registered in Tax-Heavens from State Aid*, REUTERS (Apr. 20, 2020, 9:43 AM), https://www.reuters.com/article/us-health-coronavirus-denmark-idUKKBN2221V8 [https://perma.cc/YYG2-2NMG].

553. The European Union is currently navigating these trade-offs as it seeks to construct several sectoral data pools to facilitate data sharing within Europe. *See supra* note 483 and accompanying text.

governance mechanisms that are resistant to corporate capture and adhere to fundamental administrative principles of transparency, participation, reason-giving, and review to ensure accountability. Developing publicly supported data sharing infrastructures with commensurate governance mechanisms strikes us as the most promising short-term intervention to address data inequality.

E. Developing Collective Data Governance

Throughout this Part, we have alluded to collective governance over data as illustrations of initiatives for pooling data for common benefit, empowering individuals supplying data, and as possible venues for leveraging collective bargaining power. Elinor Ostrom's work on knowledge commons and the governance practices of open source software communities provide valuable frames and models, but collective governance over data and in particular over data infrastructures remains understudied and undertheorized.[554] An in-depth examination of such arrangements is beyond the scope of this Article. However, a few observations are worth making.

Governance over data and data infrastructures is complicated by the fact that both data and infrastructures are relational concepts.[555] As Salomé Viljoen demonstrates, data's capacity to transmit social and relational meaning is not only central to the production of economic value from data, but also "renders data production especially capable of benefitting and harming others beyond the data subject from whom data is collected."[556] These features of data explain the important limitations of extant law, which, with some exceptions, views governance of data through the lens of individual rights (e.g., property or fundamental rights).[557] Viljoen argues that, instead, the aim of data governance should be "to develop the institutional responses necessary to represent the relevant population-level interests at stake in data

554. *But see* the contributions in GOVERNING KNOWLEDGE COMMONS (Brett M. Frischmann, Michael J. Madison, & Katherine J. Strandburg eds., 2014); GOVERNING PRIVACY IN KNOWLEDGE COMMONS (Madelyn Rose Sanfilippo, Brett M. Frischmann, & Katherine J. Strandburg eds., 2021). *See also* Sylvie Delacroix & Neil D. Lawrence, *Bottom-Up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance*, 9 INT'L DATA PRIV. L. 236, 240–48 (2019) (proposing trust structures to account for human vulnerability).

555. Viljoen, *supra* note 287; Kingsbury & Maisley, *supra* note 84; *see also* discussion *supra* Section I.C.

556. Viljoen, *supra* note 287, at 583.

557. Salomé Viljoen, *Data as Property?*, PHENOMENAL WORLD (Oct. 16, 2020), https://www.phenomenalworld.org/analysis/data-as-property [https://perma.cc/PQN3-NF6E]; *see also* discussion *supra* Sections II.B–C.

production . . . securing recognition and standing to shape the purposes and conditions of data production for those with interests at stake in such choices, and thus establish the terms of legitimate mutual obligation."[558] This approach, she posits, would also provide the foundation for mandatory data collection, "as long as the purposes and the conditions of such collection are derived from legitimate forms of collective self-willing and further legitimate public ends."[559]

Although we are sympathetic to Viljoen's argument, its implementation is complicated by the relational nature of infrastructures. Like data, infrastructures both shape and are shaped by relations. Data infrastructures are products of organizational dynamics of corporate forms, as their components are built and linked up through technical and legal (often contractual) relationships between designers and manufacturers of devices, software engineers, logistics personnel, and users.[560] Data infrastructures thus implicate different publics whose interests will not be always aligned.[561] The transnationality of data infrastructures also means that the relevant publics are dispersed, often unaware of, and not exposed to, each other's existence. This might explain why the emergence of data collectives thus far has been very local, with relatively clearly defined publics. Layered on top are gendered, racial, cultural, and other socio-economic dimensions that can suppress participation and exclude certain publics.[562] Legal regimes and institutions can both facilitate and foreclose connections between and among publics.[563]

---

558. Viljoen, *supra* note 287, at 64.

559. *Id.* at 72.

560. *See* discussion *supra* Section I.B.

561. *See generally* Kingsbury & Maisley, *supra* note 84.

562. This is an increasingly recognized problem in open-source software communities. *See, e.g.*, Caroline Sinders, *Designing for Community Health and Safe Spaces: The History of the JS Confs and Fighting Harassment to Maintain Healthy Open Source Communities*, CONVOCATION DESIGN + RESEARCH, FORD FOUNDATION (Nov. 23, 2020).

563. Ride sharing companies Uber and Ola had opposed drivers' requests for access to data companies have about them for purposes of developing a trade union data trust on the grounds that doing so would violate data protection rights of customers. *Uber and Ola Cabs in Legal Bid to Curtail Worker Digital Rights and Suppress Union Organised Data Trusts*, ADCU (Dec. 16, 2020), https://www.adcu.org.uk/news-posts/uber-and-ola-cabs-in-legal-bid-to-curtail-worker-digital-rights-and-supress-union-organised-data-trusts [https://perma.cc/R8NM-G747]. In 2021, the Court of Amsterdam ordered Uber and Ola to provide their drivers with greater access to data held on them and rejected both firms' claims that drivers were abusing their data access rights by seeking to use it for collective bargaining purposes. Rb. Amsterdam 11 maart 2021, Prg. 2021/134 m. nt. RDVR (Petitioners/Uber B.V.) (Neth).; Rb. Amsterdam 11 maart 2021, JBP 2021/72 m. nt. NWVG (Petitioners/Ola Netherlands B.V.) (Neth).

Despite these challenges, exploring ways in which individuals, communities, and other groupings of constituencies can be empowered to participate in datafication decisions—as well as in the design, management, and oversight of data infrastructures—is a worthwhile endeavor. Deploying technological means to create spaces for public deliberation, creating legal spaces for political and social organizing, and engaging the public in an auditing process of companies exercising significant infrastructural control over data are just some of the steps that might be taken towards remedying data inequality.

CONCLUSION

Data is more than an economic resource. Data is a medium through which economic, social, and political life is increasingly being ordered and reordered. Sufficient and sustainable access to data is unattainable for developing economies that lack the prerequisite data infrastructures necessary to generate, store, and process data on their own terms. Individuals, communities, and societies are being deprived of capabilities and possibilities to chart their own digital destinies when those that control the means of data production also control the ability to define, classify, shape, make visible or erase identities and environments, labor and leisure, conflicts and solidarities, freedoms, and oppression.[564]

Faced with this scenario, the intuition of lawyers and regulators is often to search for an appropriate legal intervention, usually guided by existing legal and regulatory frameworks and institutions. But law is not exogenous; it does not simply act (or not) *on* data. Law co-constitutes, shapes, enables, and symbiotically intertwines with data infrastructures. Extant legal paradigms and institutions may target discrete issues more or less effectively, but broader impacts of datafication, including those on individual welfare, developmental freedom, and democratic governance, are too rarely considered. When law and lawyers overlook where control over data and its constituting infrastructures is being exercised and where and why it is being entrenched, the successful contestation of outsized power to datafy becomes elusive, thereby exacerbating data inequality. As we have argued throughout this Article, effective interventions need to be attuned not only to legal but also to other infrastructural dimensions, including the politics of data infrastructures, and may need to creatively explore new regimes and institutions.

564. See generally the seminal work of AMARTYA SEN, DEVELOPMENT AS FREEDOM (1999).

There is neither a single nor an easy "fix" to the problem of data inequality. Indeed, revealing and unraveling the root causes of data inequality may take time. It may be worth pausing the pace of "leapfrogging" to (re)evaluate the degree to which contemporary patterns of datafication and "free flow" of data resemble patterns of colonial extractivism.[565] Such (re)evaluation may lead to the reimagination of existing legal domains and the realignment of legal frameworks to better account for the economics and politics of data infrastructures. Still, law—even if recast—will not be a silver bullet for remedying data inequality. Data, from the moment of its conception until its exploitation, is deeply political. Thus, interventions to remedy data inequality must be situated within continuous, iterative, inclusive, and public debate. Development organizations should encourage and create spaces that foster opportunities for communities to reclaim collective governance over data. Here, a reinvigorated conception of human rights might be brought to bear by focusing predominantly not on the invocation of individual rights, but on building transnational movements across different publics and territorial boundaries to enable and support decisive and forceful action that can confront and overcome countervailing interests.[566]

States—and their publics—must be able to experiment with digital development policies without being overly-constrained by international economic law. Naturally, such experiments will not always succeed. But, as data infrastructures are being built at a fast pace and on massive scales, the moment to chart new and alternative pathways to confront data inequality is now.

565. For a comprehensive discussion of the links between contemporary data extraction and historical practices of colonial extractivism, see generally COULDRY & MEJIAS, *supra* note 95.

566. On the processes through which human rights ideas and practices developed in cosmopolitan centers are being translated into terms for local contexts, see Sally Engle Merry & Peggy Levitt, *The Vernacularization of Women's Human Rights*, *in* HUMAN RIGHTS FUTURES 213–36 (Stephen Hopgood, Jack Snyder & Leslie Vinjamuri eds., 2017). On the transformational impacts of human rights movements, see generally GRÁINNE DE BÚRCA, REFRAMING HUMAN RIGHTS IN A TURBULENT ERA (2021).