THE YALE LAW JOURNAL

SALOMÉ VILJOEN

A Relational Theory of Data Governance

ABSTRACT. Data-governance law—the legal regime that regulates how data about people is collected, processed, and used—is the subject of lively theorizing and several proposed legislative reforms. Different theories advance different legal interests in information. Some seek to reassert individual control for data subjects over the terms of their datafication, while others aim to maximize data-subject financial gain. But these proposals share a common conceptual flaw. Put simply, they miss the point of data production in a digital economy: to put people into population-based relations with one another. This relational aspect of data production drives much of the social value and harm of data collection and use in a digital economy.

This Feature advances a theoretical account of data as social relations, constituted by both legal and technical systems. It shows how data relations result in supraindividual legal interests. Properly representing and adjudicating among those interests necessitates far more public and collective (i.e., democratic) forms of governing data production. Individualist data-subject rights cannot represent, let alone address, these population-level effects.

This account offers two insights for data-governance law. First, it better reflects how and why data collection and use produce economic value as well as social harm in the digital economy. This brings the law governing data flows into line with the economic realities of how data production operates as a key input to the information economy. Second, this account offers an alternative normative argument for what makes datafication—the transformation of information about people into a commodity—wrongful. What makes datafication wrong is not (only) that it erodes the capacity for subject self-formation, but instead that it materializes unjust social relations: data relations that enact or amplify social inequality. This account indexes many of the most pressing forms of social informational harm that animate criticism of data extraction but fall outside typical accounts of informational harm. This account also offers a positive theory for socially beneficial data production. Addressing the inegalitarian harms of datafication—and developing socially beneficial alternatives—will require democratizing data social relations: moving from individual data-subject rights to more democratic institutions of data governance.

AUTHOR. Academic Fellow, Columbia Law School. Many thanks to the members of the 2020 Privacy Law Scholars Workshop, the Information Law Institute Fellows Workshop at NYU Law, and the Digital Life Initiative Fellows Group at Cornell Tech for their careful and generous comments. Additional thanks to Ashraf Ahmed, José Argueta Funes, Chinmayi Arun, Yochai Benkler, Elettra Bietti, Julie Cohen, Angelina Fisher, Jake Goldenfein, Ben Green, Lily Hu, Woodrow Hartzog, Aziz Huq, Amy Kapczynski, Duncan Kennedy, Issa Kohler-Hausmann, Michael Madison, Lee McGuigan, Lev Menand, Christopher Morten, Helen Nissenbaum, Amanda Parsons, Angie Raymond, Neil Richards, Thomas Schmidt, Katherine Strandburg, Thomas Streinz, Mark Verstraete, Ari Ezra Waldman, and Richard Wagner. An early version of this work was presented in 2018 at Indiana University's Ostrom Workshop.

FEATURE CONTENTS

INTRODUCTION

In recent years, the technology industry has been the focus of increased public distrust, civil and worker activism, and regulatory scrutiny.[1] Concerns over datafication—the transformation of information about people into a commodity—play a central role in this widespread front of curdled goodwill, popularly referred to as the "techlash."[2]

As technology firms mediate more of our daily lives and grow more economically dominant, the centrality they place on data collection raises the stakes of data-governance law—the legal regime that governs how data about people is collected, processed, and used. As data becomes an essential component of informational capital, the law regulating data production becomes central to debates regarding how—and why—to regulate informational capitalism. There is broad consensus that current data-governance law has failed to protect technology users from the harms of data extraction, in part because it cannot account

1. Facebook's Cambridge Analytica scandal marked a turning point in the press coverage and popular sentiment toward technology companies. For more on Cambridge Analytica, see, for example, *Mark Zuckerberg Testimony: Senators Question Facebook's Commitment to Privacy*, N.Y. TIMES (Apr. 10, 2018), https://www.nytimes.com/2018/04/10/us/politics/mark-zuckerberg-testimony.html [https://perma.cc/6MKF-UEER]; and Zeynep Tufekci, *Facebook's Surveillance Machine*, N.Y. TIMES (Mar. 19, 2018), https://www.nytimes.com/2018/03/19/opinion/facebook-cambridge-analytica.html [https://perma.cc/FC9A-EJWY]. From 2015 to 2019, the number of Americans who held a positive view of technology fell by twenty-one percentage points. *See* Carroll Doherty & Jocelyn Kiley, *Americans Have Become Much Less Positive About Tech Companies' Impact on the U.S.*, PEW RSCH. (July 29, 2019), https://www.pewresearch.org/fact-tank/2019/07/29/americans-have-become-much-less-positive-about-tech-companies-impact-on-the-u-s [https://perma.cc/JA9T-J78F]. Worker activism at tech companies has increased sharply since 2016, particularly in response to contracts between technology companies and the U.S. Department of Defense and U.S. Immigration and Customs Enforcement (ICE). *See, e.g.*, #NOTECHFORICE, https://notechforice.com [https://perma.cc/TR89-N8U8]; *Worker Power in the Tech Industry*, TECH WORKERS COAL., https://techworkerscoalition.org [https://perma.cc/5CRC-7PAP]; Jimmy Wu, *Optimize What?*, COMMUNE (Mar. 15, 2019), https://communemag.com/optimize-what [https://perma.cc/F5BJ-6HXR]; Drew Harwell, *Google to Drop Pentagon AI Contract After Employee Objections to the 'Business of War,'* WASH. POST (June 1, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/06/01/google-to-drop-pentagon-ai-contract-after-employees-called-it-the-business-of-war [https://perma.cc/GZV5-FM3G].

2. The origin of the term "techlash" is commonly attributed to its use in *The Economist* in 2013. Adrian Wooldridge, *The Coming Tech-Lash*, ECONOMIST (Nov. 18, 2013), https://www.economist.com/news/2013/11/18/the-coming-tech-lash [https://perma.cc/8G7E-KDZ9]. In 2018, both the *Oxford English Dictionary* and the *Financial Times* deemed "techlash" to be a word of the year. *See Word of the Year 2018: Shortlist*, OXFORD LANGUAGES, https://languages.oup.com/word-of-the-year/2018-shortlist [https://perma.cc/M49Z-9UER]; Rana Foroohar, *Year in a Word: Techlash*, FIN. TIMES (Dec. 16, 2018), https://www.ft.com/content/76578fba-fca1-11e8-ac00-57a2a826423e [https://perma.cc/XER8-FBDQ].

for this large and growing gap between data's de jure status as the subject of consumer rights and its de facto status as quasi capital.[3]

Data-governance reform is the subject of much debate and lively theorizing, with many proposals emerging to address the status quo's inadequacy.[4] This Feature evaluates the legal conceptualizations behind these proposals—in other words, how proposed reforms conceive of what makes datafication worth regulating and whose interests in information ought to gain legal recognition. How datafication is conceptualized shapes and constrains how the law responds to datafication's effects. If data-governance law is inattentive to how data production creates social benefits and harms, it will be poorly equipped to mitigate those harms and foster data production's benefits.

This Feature's core argument is that the data-collection practices of the most powerful technology companies are aimed primarily at deriving (and producing) population-level insights regarding how data subjects relate to others, not individual insights specific to the data subject. These insights can then be applied to all individuals (not just the data subject) who share these population features.

This population-level economic motivation matters conceptually for the legal regimes that regulate the activity of data collection and use; it requires revisiting long-held notions of why individuals have a legal interest in information about them and where such interests obtain.

The status quo of data-governance law, as well as prominent proposals for its reform, approach these population-level relational effects as incidental or a byproduct of eroded individual data rights, to the extent that they recognize these effects at all. As a result, both the status quo and reform proposals suffer from a common conceptual flaw: they attempt to reduce legal interests in information to individualist claims subject to individualist remedies, which are structurally incapable of representing the interests and effects of data production's population-level aims. This in turn allows significant forms of social informational harm to go unrepresented and unaddressed in how the law governs data collection, processing, and use.

---

3. Julie E. Cohen, Between Truth and Power: The Legal Constructions of Informational Capitalism 44 (2019) ("One important byproduct of the access-for-data arrangement is a quiet revolution in the legal status of data and algorithms as (de facto if not de jure) proprietary information property."); *see also id.* at 76 (observing that there is "a growing constellation of de jure and de facto legal immunities that predominantly bolsters private economic power, that magnifies the vulnerability of ordinary citizens to manipulation, exploitation, and political disempowerment, and that threatens profound collective harm").

4. See *infra* Parts I and III for an extended discussion.

Properly representing the population-level interests that result from data production in the digital economy will require far more collective modes of ordering this productive activity.[5] The relevant task of data governance is not to reassert individual control over the terms of one's own datafication (even if this were possible) or to maximize personal gain, as leading legal approaches to data governance seek to do. Instead, the task is to develop the institutional responses necessary to represent (and adjudicate among) the relevant population-level interests at stake in data production. In other words, responding adequately to the economic imperatives and social effects of data production will require moving past proposals for individualist data-subject *rights* and toward theorizing the collective *institutional forms* required for responsible data governance.

This Feature builds on prior digital-privacy and data-governance scholarship that points out the importance of social causes and social effects of privacy erosion.[6] It takes up these insights to offer an account of *why* the social effects of privacy erosion should be considered of greater relevance — indeed, central relevance — for data-governance law. By placing data relations and their population-level effects at the center of discussions regarding why data about people is (and ought to be) legally regulated, this Feature offers two contributions to the literature on data-governance law.

First, it aligns the legal debates regarding how to govern data production with the economic transformation of data into a key input of the information economy. This in turn illuminates the growing role (and heightened stakes) of

---

5. This Feature will refer variously to the "data political economy," the "data economy," and the "digital economy." While there are distinctions between these concepts in their own right, here these all refer to sets of actors, products, business practices, and imperatives that depend on the ability to produce economic value (and political effects) through processes of data capture, transfer, and analysis. *See* MARK ANDREJEVIC, INFOGLUT: HOW TOO MUCH INFORMATION IS CHANGING THE WAY WE THINK AND KNOW 1-18, 20-21 (2013); Matthew Crain, *Financial Markets and Online Advertising: Reevaluating the Dotcom Investment Bubble*, 17 INFO., COMMC'N & SOC'Y 371, 374-81 (2014); OSCAR H. GANDY, JR., THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION 1-13 (1993); Lee McGuigan & Vincent Manzerolle, *"All the World's a Shopping Cart:" Theorizing the Political Economy of Ubiquitous Media and Markets*, 17 NEW MEDIA & SOC'Y 1830, 1831-39 (2015); Joseph Turow & Nick Couldry, *Media as Data Extraction: Towards a New Map of a Transformed Communications Field*, 68 J. COMMC'N 415, 415 (2018) (arguing that the rising *economic* importance of data extraction and analysis for digital-media companies has ushered in a "major shift" in the object of study for media and communications scholars: from the traditional focus on media content itself to how the media industries' "surveillance and population constructions" are "key infrastructural aspects of economic life").

6. *See, e.g.*, HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 1-4, 10-11 (2010); PRISCILLA M. REGAN, LEGISLATING PRIVACY: TECHNOLOGY, SOCIAL VALUES, AND PUBLIC POLICY 220-31 (1995); Julie E. Cohen, *What Privacy Is For*, 126 HARV. L. REV. 1904, 1904-06 (2013). For a more complete discussion of prior accounts, see *infra* Part I.

data-governance law as a primary legal regime regulating informational capitalism.

The descriptive contribution of this Feature details how data production in the digital economy is fundamentally relational: a basic purpose of data production as a commercial enterprise is to relate people to one another based on relevant shared population features. This produces both considerable social value and many of the pressing forms of social risk that plague the digital economy. As this Feature explores further below, data's relationality results in widespread population-level interests in data collection and use that are irreducible to individual legal interests within a given data exchange. Contending with the economic realities of data production thus expands the task of data-governance law: from disciplining against forms of interpersonal violation to also structuring the rules of economic production (and social reproduction) in the information economy.

Second, this Feature departs from prior work to offer an alternative normative account for what makes datafication wrongful. Privacy and data-governance law have traditionally governed forms of private interpersonal exchange in order to secure the benefits of data-subject *dignity* or *autonomy*. Yet as data collection and use become key productive activities (i.e., economic activities that define the contemporary economy as an information economy), new kinds of information-based harm arise. There is growing evidence of the role that digital technology plays in facilitating social and economic inequality.[7] Digital-surveillance technologies used to enhance user experience for the rich simultaneously provide

7. *See, e.g.*, VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018) (investigating the disparate impacts of sorting and monitoring technology systems on poor and working-class Americans); BEN GREEN, THE SMART ENOUGH CITY: PUTTING TECHNOLOGY IN ITS PLACE TO RECLAIM OUR URBAN FUTURE 39-116 (2019) (describing how urban technology can result in exacerbating social and political inequality); Ben Green & Salomé Viljoen, *Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought*, PROC. ACM CONF. FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 19, 20, 21-23 (2020) (observing how algorithmic formalism can entrench adverse social conditions, discrimination, and inequality); Frank Pasquale, *Two Narratives of Platform Capitalism*, 35 YALE L. & POL'Y REV. 309, 310-17 (2016) (advancing counternarratives of platform capitalism that suggest platforms can entrench inequalities, increase discrimination, undermine economic growth, and limit user agency); Neil Irwin, *To Understand Rising Inequality, Consider the Janitors at Two Top Companies, Then and Now*, N.Y. TIMES (Sept. 3, 2017), https://www.nytimes.com/2017/09/03/upshot/to-understand-rising-inequality-consider-the-janitors-at-two-top-companies-then-and-now.html [https://perma.cc/64ZF-KTSC]; Miriam Pawel, *You Call It the Gig Economy. California Calls It "Feudalism,"* N.Y. TIMES (Sept. 12, 2019), https://www.nytimes.com/2019/09/12/opinion/california-gig-economy-bill-ab5.html [https://perma.cc/E4WR-RZH5]. Other arguments highlight how the negative effects of surveillance are apportioned along lines of privilege. *See* Frank Pasquale, *Paradoxes of Privacy in an Era of Asymmetrical Social Control*, *in* BIG DATA, CRIME AND SOCIAL CONTROL 31, 31 (Aleš Završnik ed.,

methods of discipline and punishment for the poor. Algorithmic systems may reproduce or amplify sex and race discrimination.[8] Even seemingly innocuous data collection may be used in service of domination and oppression.[9] The pursuit of user attention and uninterrupted access to data flows amplifies forms of identitarian polarization, aggression, and even violence.[10] Such evidence suggests that social processes of datafication not only produce violations of personal dignity or autonomy, but also enact or amplify social inequality.

Prior accounts rightly identify the deep entanglement between the challenges of protecting autonomy in the digital economy and the realities of how data production operates as a social process: without securing better social conditions for data production for everyone, the personal benefits of robust privacy protection cannot be realized.[11] On this view, the supraindividual nature of digital-privacy

---

2018) (discussing asymmetries in surveillance and particular legal benefits afforded to the wealthy); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677-92 (2016) (identifying mechanisms by which data mining can have discriminatory impacts on protected classes); Paul Blest, *ICE Is Using Location Data from Games and Apps to Track and Arrest Immigrants, Report Says*, VICE NEWS (Feb. 7, 2020), https://www.vice.com/en/article/v7479m/ice-is-using-location-data-from-games-and-apps-to-track-and-arrest-immigrants-report-says [https://perma.cc/XB7V-3B7G].

8. *See, e.g.*, Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 1, 1-2, 5-10 (2018); Safiya Umoja Noble, *Google Search: Hyper-Visibility as a Means of Rendering Black Women and Girls Invisible*, INVISIBLE CULTURE (Oct. 13, 2013), https://ivc.lib.rochester.edu/google-search-hyper-visibility-as-a-means-of-rendering-black-women-and-girls-invisible [https://perma.cc/FWJ6-KXNL]; Ben Green, *The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness*, PROC. ACM CONF. FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 594, 596-600 (2020).

9. *See* Blest, *supra* note 7; Joseph Cox, *How the U.S. Military Buys Location Data from Ordinary Apps*, VICE NEWS (Nov. 16, 2020), https://www.vice.com/en/article/jgqm5x/us-military-location-data-xmode-locate-x [https://perma.cc/WNR6-A7PL] (detailing how the U.S military buys location data from many sources, including a Muslim prayer app with over ninety-eight million downloads).

10. *See, e.g.*, *About Us*, MEDIA MANIPULATION CASEBOOK, https://mediamanipulation.org/about-us [https://perma.cc/U7H7-88DQ]; *Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech*, DATA & SOC'Y (Oct. 17, 2018), https://datasociety.net/library/weaponizing-the-digital-influence-machine [https://perma.cc/BT7F-Q59B]; Ronan Farrow, *A Pennsylvania Mother's Path to Insurrection*, NEW YORKER (Feb. 1, 2021), https://www.newyorker.com/news/news-desk/a-pennsylvania-mothers-path-to-insurrection-capitol-riot [https://perma.cc/6MGF-FPCB]; Chinmayi Arun, *On WhatsApp, Rumours, and Lynchings*, 54 ECON. & POL. WKLY. 30, 30-33 (Feb. 9, 2019).

11. For more on the extended discussion of the democratic values at issue in data production, see Amy Kapczynski, *The Law of Informational Capitalism*, 129 YALE L.J. 1276 (2020) (reviewing SHOSHANA ZUBOFF, AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019)); and COHEN, *supra* note 3. *See* Evgeny Morozov, *Digital Socialism?*, 116/117 NEW LEFT REV. (2019); Ben Tarnoff & Moira Weigel, *Why Silicon Valley*

erosion matters because it raises additional complications for securing the benefits of robust digital-privacy protection for individuals.

This Feature departs from such accounts in that it places the inegalitarian effects of data extraction on equal footing with its autonomy-eroding effects. Privacy erosion's social effects *do* implicate the personal (and social) value of individual autonomy. But the inequality that results from data production should be considered relevant to the task of data governance for its own sake, and not only for the effects inequality has on data subjects' individual capacities for self-formation and self-enactment. This Feature thus argues that, alongside traditional concerns over individual autonomy, the social inequalities that result from data production are also forms of informational harm.

Both current and proposed data-governance law fail to adequately grasp the socioeconomic and normative centrality of data relations. This poses two problems. The first problem is conceptual: a central economic imperative that drives data production goes unrepresented in both existing and proposed laws governing datafication. As a practical matter, this leaves the law out of step with many of the ways that information creates social value and allows material forms of social informational harm to persist unaddressed. This presents U.S. data-governance law with a *sociality problem*: how can data-governance law account for data production's social effects?

The second problem is a matter of institutional design. Individualist theories of informational interests result in legal proposals that advance a range of new rights and duties with respect to information but practically fall back on individuals to adjudicate between legitimate and illegitimate information production. This not only leaves certain social informational harms unrepresented (let alone addressed), but also risks foreclosing socially beneficial information production. This presents U.S. data-governance law with a *legitimacy problem*: how can the legal regimes governing data production distinguish legitimate from illegitimate data use without relying on individual notice and choice?

The sociality problem demonstrates the need in data-governance law for an expanded account of the interests at stake in information production, while the legitimacy problem points to the need for data-governance law to expand its remit by considering whose interests are relevant for deciding whether a particular instance of data production is legitimate, and on what grounds.

This Feature offers a response to these conceptual and institutional design problems. Conceptually, it offers an account of the sociality problem that recognizes the ubiquity and the relevance of the population-level interests that result from data production. From such recognition follows this Feature's response to

Can't Fix Itself, GUARDIAN (May 3, 2018), https://www.theguardian.com/news/2018/may/03/why-silicon-valley-cant-fix-itself-tech-humanism [https://perma.cc/T6PD-QPRJ].

the legitimacy problem, which argues for governing many types of data as a collective resource that necessitates far more democratic, as opposed to personal, forms of institutional governance.

This in turn leads to a different line of inquiry regarding the legal challenges facing data-governance law. Current debates center on how to secure greater data-subject control, more robust protections for data-subject dignity, or better legal expressions of data-subject autonomy. An account of data social relations focuses future inquiry on how to balance the overlapping and at times competing interests that comprise the population-level effects of data production. This line of inquiry raises core questions of *democratic governance*: how to grant people a say in the social processes of their mutual formation; how to balance fair recognition with special concern for certain minority interests; what level of civic life achieves the appropriate level of pooled interest; and how to recognize that data production produces winners and losers and, in turn, develop fair institutional responses to these effects.

This Feature proceeds in four Parts. Part I describes the stakes and the status quo of data governance. It begins by documenting the significance of data processing for the digital economy. It then evaluates how the predominant legal regimes that govern data collection and use — contract and privacy law — code data as an individual medium. This conceptualization is referred to throughout the Feature as "data as individual medium" (DIM). DIM regimes apprehend data's capacity to cause individual harm as the legally relevant feature of datafication; from this theory of harm follows the tendency of DIM regimes to subject data to private individual ordering.

Part II presents the Feature's core argument regarding the incentives and implications of data social relations within the data political economy. Data's capacity to transmit social and relational meaning renders data production especially capable of benefitting and harming others beyond the data subject from whom the data is collected. It also results in population-level interests in data production that are not reducible to the individual interests that generally feature in data governance. Thus, data's relationality presents a conceptual challenge for data governance reform.

Part III evaluates two prominent sets of legal reform proposals that have emerged in response to concerns over datafication. Data has been extensively analogized, and proposals for reform locate data at different points on the continuum from "object-like" to "person-like."[12] On one end of this spectrum,

12. Data has been extensively analogized to both objects and aspects of personhood, spawning a robust literature on the purposes, limits, and effects of data metaphors. *See* Luke Stark & Anna Lauren Hoffmann, *Data Is the New What? Popular Metaphors & Professional Ethics in Emerging*

propertarian proposals respond to growing wealth inequality in the data economy by formalizing individual propertarian rights over data. These reforms call for formalizing an alienable right to data as labor or property, to be bought and sold in a market for goods or labor. On the other end, dignitarian reforms conceive of data as an extension of data-subject selfhood. Dignitarian reforms respond to how excessive data extraction can erode individual autonomy by strengthening the fundamental rights data subjects enjoy over their data as an extension of their personal selfhood. While propertarian and dignitarian proposals differ on the theories of injustice underlying datafication and accordingly provide different solutions, both resolve to individualist claims and remedies that do not represent, let alone address, the relational nature of data collection and use.

Finally, Part IV proposes an alternative approach: data as a democratic medium (DDM). This alternative conceptual approach recognizes data's capacity to cause social harm as a fundamentally relevant feature of datafication. This leads to a commitment to collective institutional forms of ordering. Conceiving of data as a collective resource subject to democratic ordering accounts for the importance of population-based relationality in the digital economy. This recognizes a greater number of relevant interests in data production. DDM responds not only to salient forms of injustice identified by other data-governance reforms, but also to significant forms of injustice missed by individualist accounts. In doing so, DDM also provides a theory of data governance from which to defend forms of socially beneficial data production that individualist accounts may foreclose. Part IV concludes by outlining some examples of what regimes that conceive of data as democratic could look like in practice.

Data Culture, 4 J. CULTURAL ANALYTICS 1, 5-13 (2019); Rowan Wilken, *An Exploratory Comparative Analysis of the Use of Metaphors in Writing on the Internet and Mobile Phones*, 23 SOC. SEMIOTICS 632, 635-41 (2013); Dawn Nafus, *Stuck Data, Dead Data, and Disloyal Data: The Stops and Starts in Making Numbers into Social Practices*, 15 DISTINKTION: J. SOC. THEORY 208, 208-11 (2014); Cornelius Puschmann & Jean Burgess, *Metaphors of Big Data*, 8 INT'L J. COMMC'N 1690, 1697-1701 (2014); Deborah Lupton, *Swimming or Drowning in the Data Ocean? Thoughts on the Metaphors of Big Data*, SOC. LIFE (Oct. 29, 2013), https://simplysociology.wordpress.com/2013/10/29/swimming-or-drowning-in-the-data-ocean-thoughts-on-the-metaphors-of-big-data [https://perma.cc/26BN-MJ5K]; Sara M. Watson, *Data Is the New "___,"* DIS MAG. (May 28, 2016), http://dismagazine.com/discussion/73298/sara-m-watson-metaphors-of-big-data [https://perma.cc/A44E-J7U5]; Kailash Awati & Simon Buckingham Shum, *Big Data Metaphors We Live by*, TOWARDS DATA SCI. (May 14, 2015), https://towardsdatascience.com/big-data-metaphors-we-live-by-98d3fa44ebf8 [https://perma.cc/6Q4K-KY3S]; Cory Doctorow, *Personal Data Is as Hot as Nuclear Waste*, GUARDIAN (Jan. 15, 2008), https://www.theguardian.com/technology/2008/jan/15/data.security [https://perma.cc/D34R-GAFK]; Lilly Irani, *Justice for "Data Janitors,"* PUB. BOOKS (Jan. 15, 2015), https://www.publicbooks.org/justice-for-data-janitors [https://perma.cc/7QMG-PVKX].

Before continuing, three definitional and stylistic notes regarding this Feature's use of key terms are in order:

- *Data.* For the sake of brevity, "data" refers to data about people unless otherwise noted. Data about people is the data collected as people "invest, work, operate businesses, socialize," and otherwise go about their lives.[13] This data is of greatest interest to competing digital-technology companies and to observers of the business models built from data collection. It is also deliberately more expansive than U.S. definitions of "personal data" or the closely related term "personally identifiable information."[14] Furthermore, this Feature will refer to "data" as a singular, not a plural noun. This stylistic choice is in line with the common rather than the strictly correct usage.
- *Data subject and data collector.* This Feature will use the term "data subject" to refer to the individual from whom data is being collected—often also referred to in technology communities as the "user." "Data processor" is used synonymously with "data collector" to refer to the entity or set of entities that collect, analyze, process, and use data. The definitions of "data subject" and "data processor"

13. COHEN, *supra* note 3, at 38.

14. U.S. privacy law is a patchwork of state and federal laws, several of which are discussed in greater depth in Part I. Definitions of personal data vary by regulation, but a hallmark of U.S. privacy laws is that many of the obligations they place on regulated entities are tied to "personal data" or "personally identifiable information," however defined. Some of these definitions are quite broad and encompass much, if not quite all, of the social data discussed in this Feature. For instance, the National Institute of Standards and Technology (NIST) defines personally identifiable information in the federal-agency context as "any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information." Erika McCallister, Tim Grance & Karen Scarfone, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*, NAT'L INST. STANDARDS & TECH. 2-1 (2010), https://nvpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf [https://perma.cc/6RVU-QPG4] (quoting U.S. GOV'T ACCOUNTABILITY OFF., GAO-08-536, PRIVACY: ALTERNATIVES EXIST FOR ENHANCING PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION 1, 29 (2008), https://www.gao.gov/assets/gao-08-536.pdf [https://perma.cc/H2VZ-Z8Y9]). State breach-notification laws and data-security laws typically define personal data more narrowly, focusing on sensitive categories of information like social-security numbers, credit-card and financial-account numbers, personal health data, financial data, creditworthiness data, and biometric data. For a list of state data-breach-notification laws, see *Security Breach Notification Laws*, NAT'L CONF. STATE LEGISLATURES (Apr. 15, 2021), https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx [https://perma.cc/CWU6-CMRU].

are loosely derived from the European Union's General Data Protection Regulation (GDPR).[15] While the GDPR's definition of personal data offers some capacity for nonindividualistic interpretation, any reference to "data subject" in this Feature will refer to the individual from whom or about whom data is being collected.

- *Informational Harm. Individual* informational harm refers to harm that a data subject may incur from how information about them is collected, processed, or used. In contrast, *social* informational harm refers to harms that third-party individuals may incur when information about a data subject is collected, processed, or used.

I. DATA GOVERNANCE: THE STAKES AND THE STATUS QUO

This Part describes the stakes and the status quo of data governance. It begins by documenting the significance of data production for the digital economy. It then evaluates how the predominant legal regimes that govern data collection and use — contract and privacy law — code data as an individual medium.

A. Data as an Essential Feature of Informational Capitalism

Data plays a central role in both descriptive and critical accounts that characterize the contemporary digital political economy as informational capitalism.[16]

---

15. Article 4 offers the following definition: "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." Council Regulation 2016/679, art. 4, 2016 O.J. (L 119) 1, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [https://perma.cc/2RZ3-KZKT].

16. Informational capitalism, also called surveillance capitalism and data capitalism, refers to a mode of production *centrally oriented* around extracting and processing information in order to extract and amass wealth. This transforms information — particularly information in the machine-legible form of data — into a key productive resource. *See* COHEN, *supra* note 3, at 6 ("In a regime of informational capitalism, market actors use knowledge, culture, and networked information technologies as means of extracting and appropriating surplus value, including consumer surplus."). Manuel Castells defines informational capitalism as the alignment of capitalism as a mode of production with informationalism as a mode of development. MANUEL CASTELLS, 1 THE INFORMATION AGE: ECONOMY, SOCIETY, AND CULTURE: THE RISE OF THE NETWORK SOCIETY 14-16 (2d ed. 2000) (describing the new social structure associated with the emergence of informationalism); *see also* DAN SCHILLER, HOW TO THINK ABOUT INFORMATION, at xiv, 3-4 (2007) (analyzing the transition into informationalized capitalism and

Among competing technology companies, greater access to high-quality data is a key competitive advantage that allows them to build better algorithmic products, gain better insights into their customers (or the audiences their customers want to reach), and price goods, services, or bids more advantageously.[17]

Companies engaged in data collection thus view data production as a key feature of what makes the digital economy profitable. Data about people produces revenue in three ways: companies can sell it directly, use it to improve services, or use it to predict, change, or modify behavior.[18] Of these three options, behavioral use represents by far the biggest source of revenue for technology

---

the effects of this transition for information as a commodity and a subject of theoretical inquiry). For an early discussion of these concepts, see Kevin Robins & Frank Webster, *Cybernetic Capitalism: Information, Technology, Everyday Life*, *in* THE POLITICAL ECONOMY OF INFORMATION 44, 57-70 (Vincent Mosco & Janet Wasko eds., 1988), which provides a prescient analysis of what today is called informational capitalism.

17. ERIC A. POSNER & E. GLEN WEYL, RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY 208-09 (2018). Business-facing publications emphasize the importance of data for maintaining and achieving competitive advantage. *See, e.g.*, Andrei Hagiu & Julian Wright, *When Data Creates Competitive Advantage and When It Doesn't*, HARV. BUS. REV., Jan.-Feb. 2020, at 94; Nitin Seth, *Analytics Are a Source of Competitive Advantage, If Used Properly*, FORBES (July 18, 2018, 7:15 AM EDT), https://www.forbes.com/sites/forbestechcouncil/2018/07/18/analytics-are-a-source-of-competitive-advantage-if-used-properly/?sh=50e6961d1894 [https://perma.cc/X7XJ-UTZB]. Antitrust scholars are paying increasing attention to the competitive effects of mass data collection and the locked-in advantages greater data access offers incumbent computing technologies. *See, e.g.*, MAURICE E. STUCKE & ALLEN P. GRUNES, BIG DATA AND COMPETITION POLICY 36-50 (2016); Dina Srinivasan, *Why Google Dominates Advertising Markets: Competition Should Lean on the Principles of Financial Market Regulation*, 24 STAN. L. REV 55 *passim* (2020). The near-monopolistic control of data flows by certain entities, and the competitive advantage this creates, have attracted growing regulatory attention in the European Union. *See Antitrust: Commission Launches Sector Inquiry into the Consumer Internet of Things (IoT)*, EUR. COMM'N (July 16, 2020), https://ec. europa.eu/commission/presscorner/detail/en/IP_20_1326 [https://perma.cc/8B89-C35S]. Yet, as commentators have noted, these antitrust responses may be limited if they are not accompanied by data-governance reform that attends to the economic significance of data. *See* James Surowiecki, *What Does Breaking up Big Tech Really Mean?*, MIT TECH. REV. (June 30, 2021), https://www.technologyreview.com/2021/06/30/1026502/big-tech-breakup-monopoly-antitrust [https://perma.cc/NCC3-WHCQ] ("[I]f the new antitrust movement really wants to change the digital economy, challenging the Big Four's various sketchy practices is not going to be enough. These companies' greatest competitive advantage isn't the legally dubious stuff they're doing—it's their perfectly legal access to enormous amounts of detailed and granular user data. That data helps them understand their users better than anyone else and make continuous improvements to their products and services—which in turn helps them keep their current users and add new ones, which gives them access to more data, and so on. It is the key to their growth.").

18. David Stein, Presentation at the Privacy Research Group, NYU Law School (Feb. 26, 2020); David Stein, Presentation at the Information Law Institute, NYU Law School (July 15, 2020); Email from David Stein to Salomé Viljoen (Mar. 8, 2020) (on file with author).

companies.[19] Based on available evidence, the vast majority of this revenue comes from the ad-tech industry—the business of buying and selling user attention.[20] In 2019, Google reported $134.81 billion in advertising revenue out of $160.74 billion in total revenue.[21] In the first quarter of 2020, Facebook's total

19. High-quality objective and publicly available estimates of the value of global data flows (and the source of that value in how the data is used) are difficult to obtain, and standardizing such measures is a subject of ongoing effort. *See* Kean Birch, DT Cochrane & Callum Ward, *Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech*, 8 BIG DATA & SOC'Y 1 (2021) (elaborating on how data monopolists measure, govern, and account for the value of data); Ben Williamson, *Nudging Assets*, CODE ACTS EDUC. (Sept 17, 2021), https://codeactsineducation.wordpress.com/2021/09/17/nudging-assets [https://perma.cc/W8J7-MJM3]; *A Roadmap Toward a Common Framework for Measuring the Digital Economy: Report for the G20 Digital Economy Task Force*, OECD 7-10 (2020), https://www.oecd.org/sti/roadmap-toward-a-common-framework-for-measuring-the-digital-economy.pdf [https://perma.cc/GQZ2-MAQC]; *Measurement Issues in the Digital Economy*, ECON. STAT. CTR. EXCELLENCE, https://www.escoe.ac.uk/projects/measurement-issues-in-the-digital-economy [https://perma.cc/7JGH-ZF5E]; David Nguyen & Marta Paczos, *Measuring the Economic Value of Data and Cross-Border Data Flows: A Business Perspective*, OECD 5-6 (2020), https://www.oecd-ilibrary.org/science-and-technology/measuring-the-economic-value-of-data-and-cross-border-data-flows_6345995e-en [https://perma.cc/8B4T-248M]; Diane Coyle & David Nguyen, *Cloud Computing, Cross-Border Data Flows and New Challenges for Measurement in Economics*, 249 NAT'L INST. ECON. REV. R30, R30 (2019). Some evidence pegs the global data-brokerage industry at about $200 billion annually. David Lazarus, Column, *Shadowy Data Brokers Make the Most of Their Invisibility Cloak*, L.A. TIMES (Nov. 5, 2019, 5 AM PT), https://www.latimes.com/business/story/2019-11-05/column-data-brokers [https://perma.cc/9DQ4-UNW6]. However, a significant amount of the data being bought and sold via data brokers is not data about people.

20. The use of behavioral data to improve pricing and bidding strategies in online stores or advertising auction exchanges, with the aim of capturing a greater proportion of surplus value, is a lively topic of research among the data-science and algorithmic-mechanism-design research communities and in the industry of programmatic advertising. *See, e.g.*, Hal R. Varian, *Computer Mediated Transactions*, 100 AM. ECON. REV. 1, 4-5 (2010); Liran Einav & Jonathan Levin, *Economics in the Age of Big Data*, 346 SCIENCE 1243089-1, 1243089-4 (2014); Joseph Y. Halpern & Rafael Pass, *Algorithmic Rationality: Game Theory with Costly Computation*, 156 J. ECON. THEORY 246 (2015); Eric Sodomka, Rsch. Scientist, Facebook, On How Machine Learning and Auction Theory Power Facebook Advertising, Address to the Simons Institute for the Theory of Computing (Nov. 17, 2015), https://simons.berkeley.edu/talks/eric-sodomka-2015-11-17 [https://perma.cc/GXZ6-D5RT]; Tuomas Sandholm, *Automated Mechanism Design: A New Application Area for Search Algorithms*, *in* PRINCIPLES AND PRACTICE OF CONSTRAINT PROGRAMMING–CP 2003, at 19 (Francesca Rossi ed., 2003). For a legal treatment, see Srinivasan, *supra* note 17.

21. *Annual Revenue of Google from 2002 to 2020*, STATISTA, https://www.statista.com/statistics/266206/googles-annual-global-revenue [https://perma.cc/T3JL-RHFY]; *Advertising Revenue of Google from 2001 to 2020*, STATISTA, https://www.statista.com/statistics/266249/advertising-revenue-of-google [https://perma.cc/29L6-AZJQ].

advertising revenue amounted to $17.44 billion, compared to $297 million in revenue from other streams.[22]

Advertising techniques developed to predict or to influence behavior are increasingly gaining purchase in other industries. The same capabilities that help digital companies know (or claim to know)[23] what attributes make someone likely to buy an advertised product, or that are leveraged to increase a desired behavior, can be used for other tasks. For instance, these techniques may be used to identify potential voters likely to engage on an issue or with a candidate, to identify what activities are associated with risky or risk-averse financial or health behavior, or to predict how much different people are willing to pay for the same product. These uses point toward new avenues of growth for the data economy: in political-consulting services, health insurance, financial services, and hiring.[24] Overall, the digital economy powered by these behavioral techniques represents roughly $2.1 trillion, making it the fourth-largest industry in the United States.[25]

---

22. *Facebook's Global Revenue as of 2nd Quarter 2021, by Segment*, STATISTA, https://www.statista.com/statistics/277963/facebooks-quarterly-global-revenue-by-segment [https://perma.cc/BEL9-V4Y8].

23. Digital advertising has widespread issues with fraudulent claims and inflated numbers regarding what advertisers know about users that are being targeted and whether users are even being reached. Erik Sherman, *How Companies Can Deal with Rampant Digital Ad Fraud*, FORBES (Apr. 23, 2021), https://www.forbes.com/sites/zengernews/2021/04/23/how-companies-can-deal-with-rampant-digital-ad-fraud [https://perma.cc/KV9X-XFF9]. For a fascinating account of the digital-advertising industry that explores its technological and quasi-scientific history, see Lee McGuigan, *Automating the Audience Commodity: The Unacknowledged Ancestry of Programmatic Advertising*, 21 NEW MEDIA & SOC'Y 2366 (2019).

24. Core digital-services providers often branch into other sectors. *See, e.g.*, Jay Peters, *Verily, Google's Health-Focused Sister Company, Is Getting into Insurance*, VERGE (Aug. 25, 2020), https://www.theverge.com/2020/8/25/21401124/alphabet-verily-insurance-coefficient-stop-loss [https://perma.cc/Y3NF-6RPB] (describing the launch of a new insurance subsidiary by Verily Life Sciences, an Alphabet-owned company focused on health); Brittany Straughn, *Amazon Halo Now Available for John Hancock Vitality Members*, JOHN HANCOCK (Dec. 14, 2020), https://www.johnhancock.com/about-us/news/john-hancock-insurance/2020/12/amazon-halo-now-available-for-john-hancock-vitality-members.html [https://perma.cc/9RSQ-KE6J] (announcing John Hancock Insurance's addition of Amazon's health wearable device, Halo, as a benefit to their Vitality insurance program).

25. In June 2021, the Bureau of Economic Analysis estimated the digital economy's 2019 value as roughly $2.1 trillion, placing its contribution to overall GDP at 9.6%. *Updated Digital Economy Estimates—June 2021*, BUREAU ECON. ANALYSIS, https://www.bea.gov/system/files/2021-06/DE%20June%202021%20update%20for%20web%20v3.pdf [https://perma.cc/9LDS-UCQQ]. How much of the value of the digital economy (and/or the economy overall) is due to the value of underlying data assets—and how governments or other assessors ought to value and tax such assets—is the subject of considerable and lively debate. *See* Amanda Parsons, *Tax's Digital Labor Dilemma*, 71 DUKE L.J. (forthcoming 2021) (manuscript at 16, 18),

Data's value drives consequential decisions in the digital economy. Consider just two recent examples. First, when the streaming service HBO Max (owned by WarnerMedia) launched in May 2020, it was not available on two of the largest streaming platforms, Roku and Amazon Fire TV (which together comprise sixty-three percent of viewing time in the United States) for several months.[26] The reason for this was stalled contract negotiations over access to data value: WarnerMedia wanted greater access to and control over user data and resulting advertising than either Roku or Amazon was willing to provide.[27] In order to maintain their positions regarding this data, all parties were willing to forego considerable mutual gains.[28] Second, the Trump reelection campaign's decision to partner with a small advertising-software agency called Phunware to develop its 2020 campaign app was based on the company's ability to deliver valuable electoral data:

> The Trump campaign is not paying Phunware four million dollars for an app . . . . They are paying for data. They are paying for targeted advertising services. Imagine if every time I open my phone I see a campaign message that Joe Biden's America means we're going to have war in the streets. That's the service the Trump campaign . . . ha[s] bought from Phunware. An app is just part of the package.[29]

---

https://ssrn.com/abstract=3902852 [https://perma.cc/2TAE-HUVC]. However, both technology companies and economists are keen to estimate data's value for the digital economy and recognize its significance. *See, e.g.*, *The Rise of Data Capital*, MIT TECH. REV. CUSTOM 3-4 (2016), http://files.technologyreview.com/whitepapers/MIT_Oracle+Report-The_Rise_of_Data_Capital.pdf [https://perma.cc/58C9-L2ZL] ("[F]or most companies, their data is their single biggest asset . . . . Google, Amazon, Netflix, and Uber have all realized that data is more than just a record of something that happened. Data is raw material for creating new kinds of value . . . .").

26. Julia Alexander, *Why Peacock and HBO Max Aren't on the Biggest Streaming Platforms*, VERGE, (July 15, 2020, 4:27 PM EDT), https://www.theverge.com/21324139/peacock-roku-amazon-fire-tv-hbo-max-streaming-warnermedia-nbcuniversal-disney-apple [https://perma.cc/26LM-SFBB] ("Roku commands 44 percent of viewing time in the United States, according to research released earlier this year by Conviva, and Amazon Fire TV maintains about 19 percent of viewing time.").

27. *See id.* ("The roadblock, like so many debates in the tech and media space, comes down to money and data. Essentially, both NBCUniversal (owned by Comcast) and WarnerMedia (owned by AT&T) want more control over user data and advertising generated by their apps.").

28. *See id.*

29. Sue Halpern, *How the Trump Campaign's Mobile App Is Collecting Huge Amounts of Voter Data*, NEW YORKER (Sept. 13, 2020), https://www.newyorker.com/news/campaign-chronicles/the-trump-campaigns-mobile-app-is-collecting-massive-amounts-of-voter-data [https://perma.cc/3TSS-DZXR].

Critics similarly note data production's significance for the digital economy. Jathan Sadowski identifies data as a distinct form of capital, linking the imperative to collect data to the perpetual cycle of capital accumulation.[30] Julie E. Cohen traces how platform companies like Amazon and Facebook secure quasi ownership over user data through enclosure of data and identifies the processing of information in "data refineries" as a "centrally important means of economic production."[31] In Polanyian tradition, Cohen argues that data about people represents a "fourth factor of production" that sets apart informational forms of capitalism.[32] And Shoshanna Zuboff compares data production to conquest-based forms of wealth accumulation, likening people's inner lives to a precolonial continent, invaded and strip-mined for profit by technology companies.[33] These accounts locate in datafication a particular economic process of value creation that demarcates informational capitalism from its predecessors.[34]

---

30. Jathan Sadowski, *When Data Is Capital: Datafication, Accumulation, and Extraction*, BIG DATA & SOC'Y, Jan.-June 2019, at 1, 1.

31. COHEN, *supra* note 3, at 67-68.

32. *Id.* at 47. Cohen develops her account of data's role as a factor of production in informational capitalism from the three inputs Karl Polanyi identified as basic factors of production in a capitalist political economy: land, labor, and money. The shift to industrial capitalism transformed these three inputs into commodities. Cohen argues that the subsequent shift to informational capitalism reconstitutes them again, into new datafied inputs for profit extraction. At the same time, data flows about people become a vital, fourth factor of production. *Id.* at 15-47.

33. Shoshanna Zuboff invokes colonial comparisons of invasion and forcible dispossession in her case for (1) why surveillance capitalism marks a point of departure from prior forms of capitalism as well as for (2) why surveillance capitalism results in new kinds of harm. *See* ZUBOFF, *supra* note 11, at 103-04 ("They celebrate their claim to operational spaces beyond the reach of political institutions: the twenty-first century equivalent of the 'dark continents' that drew nineteenth-century European speculators to their shores."); *id.* at 142 ("My house, my street, my neighborhood, my favorite café: each is redefined as a living tourist brochure, surveillance target, and strip mine, an object for universal inspection and commercial expropriation."); *id.* at 521 ("I say that it is not OK to have our best instincts for connection, empathy, and information exploited by a draconian quid pro quo that holds these goods hostage to the pervasive strip search of our lives."); *see also* Shoshanna Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75 (2015) (further examining the consequences of surveillance capitalism). Others more explicitly engage the comparison between data extraction and colonialism. *See, e.g.*, NICK COULDRY & ULISES A. MEJIAS, THE COSTS OF CONNECTION: HOW DATA IS COLONIZING HUMAN LIFE AND APPROPRIATING IT FOR CAPITALISM (2019).

34. This Feature will repeatedly refer to the terms "datafication" and "data extraction." Consistent with the definition above, it defines "datafication" as the transformation of information or knowledge into a commodity. It defines "data extraction" as the seamless and near-continual flow of such datafied knowledge from data subjects to data collectors (often platforms).

B. Privacy Law's Individualism

The primary regime governing the collection of such data in the United States is digital-privacy law, used here to encompass the suite of laws that together regulate how data about people is collected, processed, shared, and used.[35]

U.S. privacy law comprises an overlapping and complementary web of federal and state contract law, consumer protection, privacy torts, and sector-specific consumer rights laws. Most data collected about people is governed by contractual terms of service, subject to the Federal Trade Commission Act's Section 5 and state consumer-protection oversight.[36] A series of sector-specific privacy laws have granted additional rights to consumers over particular kinds of data, such as consumer credit data, health and financial information, and educational information. These include the Health Insurance Portability and Accountability Act (HIPAA), the Children's Online Privacy Protection Act (COPPA), the Family Educational Rights and Privacy Act (FERPA), and the Fair Credit Reporting Act (FCRA), alongside a few prominent state laws like Illinois's Biometric Information Privacy Act (BIPA) and the California Consumer Privacy Act (CCPA).[37]

---

35. *Cf.* Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. (forthcoming 2021) (manuscript at 10), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3642217 [https://perma.cc/WH8E-27JY] (describing the "dominant regime" of American privacy law as one in which corporations "are largely free to exploit data as long as they disclose their intentions in a privacy 'notice' and give consumers some 'choice' about whether they wish to share their data"). Intellectual property and trade secrecy also play a significant role in structuring current data processing. *See generally* Kapczynski, *supra* note 11, at 1515 ("The law of intellectual property and trade secrets . . . morphed to enable the capture of information and data as corporate capital, and to allow their deployment to extract surplus in new ways.").

36. *See* 15 U.S.C. § 45(a)(1) (2018) (prohibiting "unfair or deceptive acts or practices in or affecting commerce"). All states have incorporated similar consumer-protection clauses into their civil codes, and state attorney general offices use their enforcement authority under such statutes and myriad other state privacy laws to regulate consumer digital terms and services. *See* Danielle Keats Citron, *The Privacy Policymaking of State Attorneys General*, 92 NOTRE DAME L. REV. 747, 754 (2016). State attorneys general have set up specialized units or departments to bring digital privacy-related enforcement actions. *See, e.g.*, *Bureau of Internet and Tech (BIT)*, N.Y. ATT'Y GEN.'S OFF., https://ag.ny.gov/bureau/internet-bureau [https://perma.cc/L8KU-X87E]; *Privacy Unit*, CAL. ATT'Y GEN.'S OFF., https://oag.ca.gov/privacy [https://perma.cc/7FLK-AB3J]. For a list of state privacy laws, see *Privacy Laws by State*, ELEC. PRIV. INFO. CTR., https://epic.org/privacy/consumer/states.html [https://perma.cc/VFW9-3US8].

37. Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C. (2018)); Children's Online Privacy Protection Act of 1998 (COPPA), 15 U.S.C. § 6501-06 (2018); Family Educational Rights and Privacy Act of 1974 (FERPA), 20 U.S.C. § 1232g (2018); Fair Credit

Many of the sector-specific laws are based on the Fair Information Practice Principles (FIPPs).[38] FIPPs are an influential set of guidelines and recommendations; they lay out standards that typify fair-data processing, serve as a continued model from which new privacy protections and industry best practices are developed, and offer guidelines for how the Federal Trade Commission (FTC) and the other state and federal bodies tasked with enforcing privacy laws evaluate the privacy promises made by industry.[39] FIPPs equate fair-data processing with practices that grant individuals meaningful control over their data. This includes requiring users to give informed *consent* to data being processed and giving users *notice* regarding how their data is used.

In lieu of enforcing the FIPPs directly, FTC uses its general authority under the FTC Act to enforce the contractual promises companies make to data subjects. In practice, the combination of FIPPs-inspired sectoral laws and FIPPs-guided FTC enforcement results in the much-maligned privacy regime known as "notice and consent" (also referred to as "notice and choice").[40] Under this regulatory regime, the terms and conditions of digital services like search engines, social networks, mobile phone apps, and other digitally mediated services

Reporting Act of 1970 (FCRA), 15 U.S.C. § 1681 (2018); Biometric Information Privacy Act (BIPA), 740 Ill. Comp. Stat. 14/10, 14/15 (2021); California Consumer Privacy Act of 2018 (CCPA), Cal. Civ. Code §§ 1798.100-1798.199.100 (West 2021).

38. FIPPs are "[p]rinciples that are widely accepted in the United States and internationally as a general framework for privacy and that are reflected in various federal and international laws and policies." *Computer Security Resource Center Glossary*, Nat. Inst. Standards & Tech., https://csrc.nist.gov/glossary/term/fipps [https://perma.cc/JTT4-JJ46].

39. FIPPs were originally named in an influential report commissioned to explore the ways in which entities use computational automated methods to collect and use personal information. *See* U.S. Dep't Health, Educ. & Welfare, DHEW Publ'n No. (OS) 73-94, Records, Computers, and the Rights of Citizens, at xx-xxi (1973).

40. This regime and its basic concepts of click-through contracts and consent as the basis for legitimate action have been given exhaustive treatment in the literature. On consent and the legal theory of legitimacy for click-through or standardized consumer contracts, see, for example, Nancy Kim, Consentability: Consent and Its Limits 91-116 (2019); Nancy Kim, Wrap Contracts 126-46 (2013); and Margaret Radin, Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law *passim* (2012). On click-through digital consent more specifically, see Elizabeth Edenberg & Meg Leta Jones, *Analyzing the Legal Roots and Moral Core of Digital Consent*, 21 New Media & Soc'y 1804, 1804-05 (2019); Woodrow Hartzog & Neil Richards, *Privacy's Trust Gap: A Review*, 126 Yale L.J. 1180, 1197-98 (2017) [hereinafter Hartzog & Richards, *Privacy's Trust Gap*] (reviewing Finn Brunton & Helen Nissenbaum, Obfuscation: A User's Guide for Privacy and protest (2015)); Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 Stan. Tech. L. Rev 431, 434 (2016) [hereinafter Richards & Hartzog, *Taking Trust Seriously*]; Andrea M. Matwyshyn, *Technoconsen(t)sus*, 85 Wash. U. L. Rev. 529 *passim* (2007); Elettra Bietti, *Consent as a Free Pass: Platform Power and the Limits of the Informational Turn*, 40 Pace L. Rev. 307, 329-31 (2020); and Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 Harv. L. Rev. 1880, 1883-85 (2013).

are presumptively valid as long as consumers are offered notice of the data being collected about them and consent to its collection.[41]

The resulting privacy-law regime conceives of data as an individual medium—it focuses legal inquiry and accords legal relevance to data's potential to cause personal harm and as therefore appropriately subject to private, individual ordering. This conceptualization of "data as an individual medium" (DIM) privileges data processing's capacity to transmit knowledge about the data subject over its capacity to transmit knowledge about others. Under DIM, this individualist knowledge transmission is the legally and normatively relevant feature of datafication.

Notice-and-consent structures the basic legal relationship between the individual consumer (the "data subject") and the digital service provider (the "data processor"). Sectoral privacy laws affirmatively grant data subjects some additional rights and impose additional duties on data processors within this relationship, but most follow a notice-and-consent template. For instance, rights to greater detail regarding data use and the duties of companies to affirmatively obtain opt-in consent (as opposed to the more passive opt-out consent) are common features of such laws.[42] Other laws grant consumers rights that strengthen certain forms of individual choice and individual control. For example, the CCPA grants data subjects rights to request information about what data is being collected about them and whether any of their personal data is being sold or disclosed to third parties.[43] It also grants data subjects the right to opt out of the

41. *See Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*, FED. TRADE COMM'N 48-60 (2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf [https://perma.cc/X335-A7Z3]; Woodrow Hartzog, *The New Price to Play: Are Passive Online Media Users Bound by Terms of Use?*, 15 COMMC'N L. & POL'Y 405, 432 (2010).

42. *See, e.g.*, Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C. (2018)). HIPAA is not only a privacy law, but its Privacy Rule limits the circumstances under which Personal Health Information (PHI) (as defined by the Act) may be used, sold, or disclosed, among other rights granted to patients. Other than as required by law or to facilitate treatment, payment, or healthcare operations, covered entities must obtain written authorization from the individual to disclose their PHI. 45 C.F.R. § 164.502 (2020); *accord, e.g.*, Children's Online Privacy Protection Act of 1998 (COPPA), 15 U.S.C. § 6501 (2018); 16 C.F.R. § 312.5(c) (2020); Family Educational Rights and Privacy Act of 1974 (FERPA), 20 U.S.C. § 1232g (2018); 34 C.F.R. § 99.30 (2020). The Gramm-Leach-Bliley Act (GLBA) only requires that entities offer consumers weaker "opt-out" rights but does require entities to provide consumers with annual notices detailing the information they collect. Pub. L. No. 106-102, § 502(b), 113 Stat. 1338, 1437 (1999); 17 C.F.R. §§ 160.1-160.9 (2020).

43. California Consumer Privacy Act of 2018 (CCPA), CAL. CIV. CODE §§ 1798.100(a)(1), 1798.110(3), 1798.135 (West 2021).

sale of their personal information.[44] The FCRA, for example, grants data subjects the right to dispute the accuracy of information in their credit reports and to have inaccurate information be updated or deleted.[45] And HIPAA grants patients the right to access their health information, to receive notice regarding how their information may be used and shared, and to consent to certain uses of their health information.[46] These laws grant consumers some additional rights but (absent a few notable but narrow exceptions) the onus remains on data subjects to exercise these rights.[47]

Existing privacy laws generally contemplate individual informational harm of the following forms.

- *Consentless collection*. Collecting data about someone without their consent is the most basic and fundamental form of informational harm contemplated by privacy laws. Obtaining personal information without consent is considered a violation of that person's right to control how information about them is used.[48] This violation harms data-subject autonomy and dignity by denying the data subject's right to informational self-determination.[49]

---

44. *See id*. *But see* Salomé Viljoen, *The Promise and Pitfalls of California's Consumer Privacy Act*, DIGIT. LIFE INITIATIVE (Apr. 11, 2020), https://www.dli.tech.cornell.edu/post/the-promise-and-pitfalls-of-the-california-consumer-privacy-act [https://perma.cc/4YJT-Q892] (canvassing the law's deficiencies).

45. Fair Credit Reporting Act of 1970 (FCRA), 15 U.S.C. § 1681i(a)(5)(A) (2018).

46. 45 C.F.R. § 164.524 (2020) (granting individuals a right to access their PHI); *id.* § 164.520 (granting individuals a right to adequate notice of the use and disclosure of their PHI); *id.* § 164.502 (requiring patient consent). Like FCRA discussed below, HIPAA also includes a few affirmative data-processing obligations and specifies certain data uses that are not subject to individual consent. However, the majority of health data-sharing contracts do not rely on these exceptions, and instead use a combination of the law's anonymity rules and patient consent to share health data. *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1736-38 (2010).

47. Certain elements of FCRA are a notable exception. For example, alongside consumer rights, FCRA places affirmative limits on who may use consumer reports for which purposes. Fair Credit Reporting Act of 1970 (FCRA), 15 U.S.C. § 1681c(a) (2018). Other exceptions include uses of personally identifiable information forbidden under HIPAA. Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191 § 1177, 110 Stat. 1936, 2023 (1996).

48. *See, e.g.*, ALAN F. WESTIN, PRIVACY AND FREEDOM: LOCATING THE VALUE IN PRIVACY 7 (1967).

49. This concept of undermining informational self-determination is closely linked to articulations of privacy as control. *See id*. at 7 (defining privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"). See *supra* Part I for further discussion of privacy as control. *See also* Cohen, *supra* note 6, at 1905 ("[Privacy] protects the situated practices of boundary management through which the capacity for self-determination develops.").

- *Sludgy consent*. A corrupted architecture or design process may result in an appearance of consent that in fact violates or undermines true consent. These may include engineering consent through design features that make opting out difficult or almost impossible or using behavioral insights to heavily influence data subjects toward granting consent.[50] Like consentless collection, sludgy consent undermines the true will of data subjects in ways that thwart their capacity for informational self-determination.
- *Harms of access*. Harms of access may occur when people are denied access to information about themselves, violating notions of informational self-determination, or when people are unable to limit or control access to information about themselves by others.[51] Harms of excessive access may include harassment and chilling effects on self-expression.[52]
- *Reidentification*. Individuals may be harmed when their identifiable personal data is released, whether intentionally or as a result of a data breach or hack. In some cases, disclosure causes immediate harm (e.g., reputational harm). Harm may also result from various inappropriate uses, including identity theft or stalking. Many privacy statutes guard against reidentification harm by allowing freer processing and use of information that has been (at least nominally)

50. *See* Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 GEO. L. TECH. REV. 1, 18-33 (2019). These design features are frequently termed "dark patterns." *See, e.g.*, Harry Brignull, *Dark Patterns: Inside the Interfaces Designed to Trick You*, VERGE (Aug. 29, 2013, 11:15 AM), https://www.theverge.com/2013/8/29/4640308/dark-patterns-inside-the-interfaces-designed-to-trick-you [https://perma.cc/6TC2-VQP8]. Such designs frequently take advantage of behavioral insights from psychology and behavioral economics that are widely used to "nudge" individuals toward socially desirable outcomes but deploy them for more socially dubious ends. *See, e.g.*, Richard H. Thaler, *Nudge, Not Sludge*, SCI. MAG., Aug. 2018, at 431.

51. In the copyright realm, Shyamkrishna Balganesh makes a similar and related claim regarding a "disseminative harm," when creators' rights to determine whether and when their works are shared have been violated, which he identifies as "compelled authorship." Shyamkrishna Balganesh, *Privative Copyright*, 73 VAND. L. REV. 1, 8-20 (2020).

52. Privacy scholars arguing for protection against online harassment and gender-based violence as privacy enhancing argue that harassment may have a chilling effect on the expressive freedoms of vulnerable groups. *See, e.g.*, Danielle Keats Citron, *Law's Expressive Value in Combating Cyber Gender Harassment*, 108 MICH. L. REV. 373, 374-78 (2009); Danielle Keats Citron & Jonathon W. Penney, *When Law Frees Us to Speak*, 87 FORDHAM L. REV. 2317, 2318-21 (2019). *See generally* SCOTT SKINNER-THOMPSON, PRIVACY AT THE MARGINS 86 (2020) (explaining how privacy offers a form of expressive resistance to surveillance regimes).

stripped of identifiers that can be used to reidentify individuals.[53] Statutes also directly address data breaches.[54]

- *Inaccuracy and discrimination.* Privacy laws also include a few thicker conceptions of individual informational harm that capture how certain forms of knowledge may cause people to lose out unfairly on important opportunities. For instance, the FCRA includes a right to accurate information and the right to delete inaccurate information in credit reports.[55] Ban-the-box initiatives similarly prohibit employers from asking about criminal convictions on employment applications, on the theory that this information may unjustly foreclose employment opportunities to deserving applicants.[56]

These forms of informational harm are individual; they identify how information flows may be produced or used in a way that may harm the data subject.

C. *Critiques of Privacy Law and Their Motivating Accounts*

While there is general scholarly agreement that data governance is in need of repair, critiques of the digital economy offer different diagnoses of *why* the status quo is insufficient, *what* the stakes of failure are, and *on what grounds* data gov-

---

53. One prominent such example is HIPAA. Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-91, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C. (2018)). Premising free processing of information on anonymization is widespread, though increasingly vexed. *See* Ohm, *supra* note 46.

54. All fifty U.S. states have passed data-breach laws that require entities affected by a data breach to notify their customers about the breach and take specific steps (that vary by statute) to remedy effects of the data breach. *See, e.g.*, N.Y. GEN. BUS. LAW §§ 899-AA to -BB (McKinney 2021) ("Notification of Unauthorized Acquisition of Private Information; Data Security Protections.").

55. Fair Credit Reporting Act of 1970 (FCRA), 15 U.S.C. §§ 1681, 1681i (2018).

56. Fourteen states have ban-the-box laws that apply to private employers. *See* Beth Avery & Han Lu, *Ban the Box: U.S. Cities, Counties, and States Adopt Fair Hiring Policies*, NAT'L EMP. L. PROJECT (Sept. 30, 2020), https://www.nelp.org/publication/ban-the-box-fair-chance-hiring-state-and-local-guide [https://perma.cc/8DBM-H84M]. However, research suggests that in jurisdictions that have passed ban-the-box laws, employers are more likely to discriminate against young Black applicants. *See* Amanda Agan & Sonja Starr, *Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*, 133 Q.J. ECON. 191, 208-11, 222 (2018); *cf.* Osborne Jackson & Bo Zhao, *The Effect of Changing Employers' Access to Criminal Histories on Ex-Offenders' Labor Market Outcomes: Evidence from the 2010-2012 Massachusetts CORI Reform* (Fed. Rsrv. Bank of Bos., Working Paper No. 16-30, 2017) (offering a "labor supply account" of the change, according to which fewer ex-offenders are hired because they become more selective about where to apply once criminal history questions are no longer a bar).

ernance fails. These diagnoses rest on different underlying claims about how information may cause harm, how information may benefit people, and how legal reformers should approach the project of data governance.

1. Traditional Accounts: Privacy as Control and Access

Much ink has been spilled on how privacy law fails to secure data-subject autonomy and thus inhibits the realization of the individual and societal goods associated with privacy. Those critiques have provided several different accounts of how and why notice-and-consent regimes fail to ensure that people retain control and access over their data. Some argue that, like many "shrinkwrap" contracts, privacy terms of service that incorporate notice-and-consent provisions operate from a legal fiction.[57] Individuals do not read the privacy policies to which they consent and have no real way to bargain over the terms they contain.[58] Others emphasize that personal data is nonrivalrous, nonextinguishable, and reusable, meaning that how it flows and how it is used can change as technologies and business models evolve. This makes data ill-suited to a regulatory approach premised on a one-time exercise of informed, individual choice.[59] Consent is easily circumvented or engineered via dark patterns, particularly in

57. "Shrinkwrap" contracts refer to the boilerplate contracts that are included as part of the packaging of the product (hence the name). Usage of the product is deemed acceptance of the contract. Another term for these take-it-or-leave-it contracts is a contract of adhesion.

58. *See* Katherine J. Strandburg, *Free Fall: The Online Market's Consumer Preference Disconnect*, 2013 U. CHI. LEGAL F. 95, 143; Jonathan A. Obar & Anne Oeldorf-Hirsch, *The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services*, 23 INFO. COMMC'N & SOC'Y 128, 140-42 (2020). Privacy policies are often long and full of legalese. They are also pervasive. One study from 2008 found that it would take an average user seventy-six days to read all the privacy policies they encountered in one year alone, with a nationwide annual estimated opportunity cost of $781 billion. *See* Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 J.L. & POL'Y FOR INFO. SOC'Y, 543, 564 (2008).

59. This shortcoming has been given exhaustive treatment by many privacy scholars. Neil Richards and Woodrow Hartzog provide a useful typology categorizing the different ways consent fails to secure privacy in the digital context. *See* Neil Richards & Woodrow Hartzog, *The Pathologies of Digital Consent*, 96 WASH. U. L. REV. 1461 (2019). Elettra Bietti provides a helpful exploration of the normative stakes of this failure. *See* Bietti, *supra* note 40. Notice and consent's blunt emphasis on one-time consent at the point of collection also aligns poorly with contextually specific concerns over appropriate information flow. *See* NISSENBAUM, *supra* note 6; *see also* Madelyn Sanfilippo, Brett Frischmann & Katherine Standburg, *Privacy as Commons: Case Evaluation Through the Governing Knowledge Commons Framework*, 8 J. INFO. POL'Y 116 (2018) (offering a "rules-in-use" concept to encompass both the nominal rules and actual practices that govern ethical information transmission).

digital settings designed for optimal data extraction.[60] Nor does notice and consent respond to the regulatory gap between the legal requirement to protect privacy and how privacy may be facilitated or eroded in practice via technical design.[61]

While many critiques of privacy law have focused on the failure of notice and consent to secure the individual and societal goods of privacy, there are also competing accounts about what, exactly, those goods are.[62] In general, legal and philosophical accounts consider privacy a predicate condition or instrumental right—part of what a just society offers in order to secure robust protection for individual autonomy or individual dignity.[63] On this view, privacy erosion threatens the vital conditions that foster the individual's ability to think for herself, enjoy a privileged relationship to her inner desires, know her own mind and express it as she chooses, and be in charge of her own formation as a social, political, and economic being.

The focus on individual selfhood is expressed in the canonical purpose of data governance: informational self-determination.[64] This purpose is consistent with the classic legal view of privacy as control, which offers ways to secure and enact self-determination. Many early and influential legal theories of privacy adopted the view of privacy as a particular form of control. For example, Alan Westin's *Privacy and Freedom: Locating the Value in Privacy* defines privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."[65] Charles Fried defines privacy as "not simply an absence of information about us in the minds of others[,] rather . . . the control we have over information about

60. *See* Solon Barocas & Helen Nissenbaum, *Big Data's End Run Around Anonymity and Consent*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 44, 56-61 (Julia Lane, Victoria Stodden, Stefan Bender & Helen Nissenbaum eds., 2014). On dark patterns, see Arvind Narayanan, Arunesh Mathur, Marshini Chetty & Mihir Kshirsagar, *Dark Patterns: Past, Present and Future*, 18 ACM QUEUE 68 (2020).

61. *See* Ari Ezra Waldman, *Privacy Law's False Promise*, 97 WASH. U. L. REV. 773, 786 (2020); WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018).

62. *See* DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 1 (2008) ("Privacy, however, is a concept in disarray. Nobody can articulate what it means. Currently, privacy is a sweeping concept, encompassing (among other things) freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and seizures.").

63. NISSENBAUM, *supra* note 6.

64. *See* Richards & Hartzog, *supra* note 35 (manuscript at 23); Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1695-96 (2020).

65. WESTIN, *supra* note 48, at 7.

ourselves."[66] More recent scholarship has similarly adopted this view. A. Michael Froomkin defines privacy as "the ability to control the acquisition or release of information about oneself."[67] Jerry Kang defines it as "an individual's control over the processing—that is, the acquisition, disclosure, and use—of personal information."[68]

Informational self-determination is also consistent with the classic legal view of privacy as access. On this view privacy is a condition, "measured in terms of the degree of access others have to you through information, attention and proximity."[69] Accounts of privacy as access are similar to accounts of privacy as control, yet distinct in that they theorize privacy not as a kind of agency that can be exercised, but a condition under which individuals sometimes find themselves. This in turn makes proponents of this view less willing to reduce normative accounts of privacy to determinations of who retains rightful control over information.[70] Ruth Gavison, a proponent of this view, traces this account in privacy laws that share a concern with intrusions of knowledge and information: under what conditions knowledge of one may be gained, what may be known by whom, how such knowledge may be used, and what effects such uses of knowledge may produce.[71]

2. *Alternative Accounts: The Social Value of Privacy*

Others link the failure of notice and consent to the social nature of privacy harm, noting the contextual nature of information flow,[72] the collective action problems and market failures it produces,[73] the externalities from individual

66. Charles Fried, *Privacy*, 77 YALE L.J. 475, 482 (1968) (emphasis omitted).

67. A. Michael Froomkin, *The Death of Privacy?*, 52 STAN. L. REV. 1461, 1463 (2000).

68. Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193, 1203 (1998).

69. NISSENBAUM, *supra* note 6, at 70.

70. *See, e.g.*, Jeffrey Raiman, *Privacy, Intimacy and Personhood*, 6 PHIL. & PUB. AFFS. 26, 30 (1976).

71. This list accords with Ruth Gavison's highly influential view of privacy as a measure of the access others have to you through information, attention, and physical proximity. *See* Ruth Gavison, *Privacy and the Limits of Law*, 89 YALE L.J. 421, 423 (1980).

72. *See* NISSENBAUM, *supra* note 6, at 127-57.

73. *See* Strandburg, *supra* note 58, at 95-97.

transactions,[74] and the epistemic constraints of individualistic frames of reference.[75] These diagnoses track alternative accounts of what privacy is for, emphasizing the social value of privacy and rejecting the atomistic conceptions behind privacy protection as informational self-determination. These accounts advance a thicker conception of autonomy that includes privacy's importance in fostering conditions of public citizenship and public governmentality.

For instance, Priscilla M. Regan argues that privacy is socially important because it facilitates democratic political flourishing through its protection of free association and free speech.[76] She also emphasizes the common stakes of privacy, given market forces that make it difficult for any one individual to have privacy unless a minimum is guaranteed to everyone.[77] Helen Nissenbaum develops an account of privacy as appropriate information flow, where context-appropriate information sharing is determined by reference to socially developed norms.[78] On this account, privacy is a claim to an appropriate information flow, governed by the "web of constraints" that make up the various norms of social life.[79] Practices that erode privacy and disturb or sunder this web of constraints "are not merely threatening" privacy as a marginal value, but "potentially tearing at the very fabric of social and political life."[80]

Julie E. Cohen offers a variant of the social privacy account that aims to depart from the liberal conception of the autonomous subject, arguing for privacy as vital for the socially constructed subject instead.[81] For this subject, "[p]rivacy shelters dynamic, emergent subjectivity" from data-driven attempts to render these subjects "fixed, transparent, and predictable."[82] This capacity is vital for self-definition, critical self-reflection, and informed citizenship—the necessary conditions for liberal democracy.[83]

---

74. *See* Solove, *supra* note 40, at 1889-93; Joel R. Reidenberg, N. Cameron Russell, Alexander J. Callen, Sophia Qasir & Thomas B. Norton, *Privacy Harms and the Effectiveness of the Notice and Choice Framework*, 11 I/S: J.L. & POL'Y FOR INFO. SOC'Y 485, 492-94 (2015); Omri Ben-Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 104 (2019).

75. *See* COHEN, *supra* note 3, at 67.

76. REGAN, *supra* note 6, at 225.

77. *See id.* at 227-31. Neil M. Richards advances a similar claim regarding the necessity of intellectual privacy for robust free expression. *See* Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 403-07 (2008).

78. NISSENBAUM, *supra* note 6, at 140-41.

79. *Id.* at 128.

80. *Id.*

81. Cohen, *supra* note 6, at 1906-11.

82. *Id.* at 1905.

83. *Id.*

Taken together, these thick accounts advance two arguments regarding the social effects of datafication (or privacy erosion). First, that strong privacy-preserving conditions to foster an individual's capacity for self-formation can only exist at a societal level; that is, for one person to have privacy (and thus the necessary conditions for self-formation), all of us must have privacy. Privacy is thus a value we must achieve societally, not individually. Second, that there is *social value* in granting an individual her capacity to develop self-knowledge and enact that knowledge (secured by strong privacy protections). For example, a flourishing democracy requires individuals who know and can enact their own will. Privacy, and the individual self-formation it fosters, are thus important not only for an individual as an individual, but also for society more generally, because we cannot have democracy without autonomously acting citizens. Both arguments emphasize the significant social effects and consequences of datafication and privacy erosion.

These thicker accounts rightly identify the social effects that drive privacy erosion and the social consequences of privacy erosion, and may indeed place central importance on the social consequences of datafication. However, both arguments ground the social benefits of addressing datafication in its deleterious effects on the ability of individuals to engage in self-knowledge formation and self-enactment. Thus, the normative basis of these arguments remains individual autonomy: datafication is wrongful, and harmful both for individuals and society, when it threatens the capacity for individuals to develop and act on their self-will. Thus, while these accounts do center the social effects of datafication, such effects serve to heighten the stakes or increase the challenges of the primary task—securing privacy protections against datafication in order to secure conditions of individual self-formation and self-enactment.

Both standard and thick accounts of autonomy inform how critics view the stakes of privacy law's failure. On these accounts, data-production practices are wrong when they lead to manipulation, erode self-determination in the data market (and beyond), chill self-expression, or involve forms of data extraction and algorithmic governmentality that infringe on an individual's capacity to act as a moral agent.[84]

---

84. Many privacy- and digital-rights activists focus on these effects, especially in the context of private systems that profit from personal violation. On manipulation, see Susser, Roessler & Nissenbaum, *supra* note 50, at 4-12, 34-44; and Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1007-12, 1024-34 (2014). On eroded self-determination, see HARTZOG, *supra* note 61, at 21-55; and Richards & Hartzog, *supra* note 59, at 1476-91. On chilling effects of self-expression, see Citron, *supra* note 52; and Citron & Penney, *supra* note 52, at 2319-20, 2329-32. On data extraction and algorithmic governmentality, see Jennifer Cobbe & Elettra Bietti, *Rethinking Digital Platforms for the Post-Covid-19 Era*, CTR. FOR INT'L GOVERNANCE INNOVATION (May 12, 2020), https://www.cigionline.org/articles/rethinking-digital-platforms-post-covid-19-era [https://perma.cc/9D8T-8RUV].

These accounts also guide views on how data governance should be reformed. As I discuss in Part III, concerns over autonomy and dignity guide dignitarian efforts to reduce datafication's commodification of inner life and to increase the regulatory oversight of data flows that act back on users in ways that wrongfully undermine their self-will. Concern over loss of control and lack of clear legal rights to data's value motivate propertarian efforts to formalize these rights for data subjects. These reforms and others focus on increasing data subjects' capacity to determine how (and under what conditions) their data is collected, processed, and used.

II. DATA RELATIONS AND THEIR SOCIAL EFFECTS

One way to evaluate different theories of data governance is to examine how such theories conceive of (and propose to act upon) the social-relations structured by data flows.[85] To understand the significance of data's relationality, let us consider with greater specificity how data relates people to one another, how such relations may produce social effects, and which of these relations are (and are not) accorded legal relevance by current and proposed forms of data-governance law.

A. Data Governance's Sociality Problem

In July 2018, privacy activists reported that the Federal Bureau of Investigation (FBI), along with the National Institute of Standards and Technology, were evaluating the effectiveness of tattoo-recognition technology.[86] To conduct this evaluation, FBI provided access to their TAG-IMAGE database — which includes images of thousands of prisoner tattoos collected from prison inmates and arrestees — to nineteen corporate and academic groups with the goal of developing

85. This Feature adapts its concept of "data relations" from prior work. Nick Couldry and Ulises A. Mejias use the term "data relations" to describe the process of capturing and processing social data, which they argue results in a new social order based on continuous tracking. Nick Couldry & Ulises A. Mejias, *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, 20 TELEVISION & NEW MEDIA 336, 336 (2018). Data social relations are constituted by both legal and technical systems that influence how data is created, collected, transmitted, and used.

86. Dave Maass, *FBI Wish List: An App that Can Recognize the Meaning of Your Tattoos*, ELEC. FRONTIER FOUND. (July 16, 2018), https://www.eff.org/deeplinks/2018/07/fbi-wants-app-can-recognize-meaning-your-tattoos [https://perma.cc/V48D-DD6V].

image-recognition technologies capable of identifying individuals by their tattoos, as well as identifying tattoos that are markers of various gang affiliations.[87]

Consider a scenario where FBI partners with a company, "TattooView AI," to provide tattoo-recognition products and automated matching — not only identifying a particular individual via their tattoo, but also determining whether their tattoo connotates gang membership more generally.[88] This tool is then used by

87. TAG-IMAGE is one of several forms of biometric markers included in the Next Generation Identification system used to automate processes of biometric identification capabilities and extend the tracking of biometric markers beyond those included in the FBI's Automated Fingerprint Identification System. The Biometric Center of Excellence (BCOE) is the agency's primary group working to develop biometrics and identity management. BCOE notes that tattoos and other biometric markers have possible uses for law enforcement well beyond purposes of identity verification. Most notably, automated recognition services allow investigators to use a probe or query image to find similar images. "While the value of image-to-image matching technology is obvious from an identification perspective, the benefits of knowing the symbolism and background behind tattoos and graffiti can be equally valuable. From an intelligence standpoint, certain symbols or graffiti may be used to help establish whether an individual is associated with a particular gang, terrorist organization, or extremist group. This may help determine the extent to which the individual or gang poses a threat to law enforcement or the community, and possibly to recognize and link crimes across the country." CJIS Link, *Image-Based Matching Technology Offers Identification and Intelligence Prospects*, FED. BUREAU INVESTIGATION (Dec. 28, 2012), https://www.fbi.gov/services/cjis/cjis-link/image-based-matching-technology-offers-identification-and-intelligence-prospects [https://perma.cc/WU7E-K4WE].

88. In addition to the 2018 National Institute of Standards and Technology-FBI trial, this scenario gains plausibility from three facts.

First, law enforcement already tracks and identifies gang membership on the basis of certain shared tattoos. On gang databases generally, see Stefano Bloch, *Are You in a Gang Database?*, N.Y. TIMES (Feb. 3, 2020), https://www.nytimes.com/2020/02/03/opinion/los-angeles-gang-database.html [https://perma.cc/6CJX-YLR3], which notes that "[s]elf-identifying as a gang member, in addition to tattoos and officers' descriptions of 'gang related' clothing, are used to make a gang distinction." On the Los Angeles Police Department's partnership with Palantir and California Databases, see Caroline Haskins, *Scars, Tattoos, and License Plates: This Is What Palantir and the LAPD Know About You*, BUZZFEED NEWS (Sept. 29, 2020, 3:00 PM ET), https://www.buzzfeednews.com/article/carolinehaskins1/training-documents-palantir-lapd [https://perma.cc/TS5S-SH6V], which asserts that "[w]ith Palantir, police can search for people by name. But . . . they can also search by race, gender, gang membership, tattoos, scars, friends, or family. 'Male, White, Peckerwood Gang, Skull Tattoo.' 'Person, Male, Hispanic, Vineland Boys, Rosary Tattoo'"; and *CalGang®: About the CalGang® Unit*, CAL. OFF. ATT'Y GEN. (2021), https://oag.ca.gov/calgang [https://perma.cc/TM8E-TXHS]. On the United Kingdom's gang database, see *Gangs Violence Matrix*, METRO. POLICE, (2021), https://www.met.police.uk/police-forces/metropolitan-police/areas/about-us/about-the-met/gangs-violence-matrix [https://perma.cc/S2W8-UAQH]. *See also* James A. Densley & David C. Pyrooz, *The Matrix in Context: Taking Stock of Police Gang Databases in London and Beyond*, 20 YOUTH JUST. 11 (2019) (assessing common critiques of gang databases). On the Georgia

law enforcement to identify potential gang members for heightened police observation.

This biometric data flows across several parties—from the initial arrestee to managers of TAG-IMAGE, to third parties such as TattooView AI, to a law enforcement officer who detains a suspected gang member. It also flows across several legal regimes: criminal law, government procurement and trade secrecy law, contract law, and privacy law. Importantly, this flow begins and ends with two human events: first, a person has his tattoo photographed and added to TAG-IMAGE (let's call him Adam); and second, a person with the same tattoo is detained using that image data (let's call him Ben).[89]

Criminal Street Gang Database, see Joshua Sharpe, *Georgia Gang Database Has Law Enforcement Hopeful, Critics Worried*, ATLANTA J.-CONST. (Feb. 5, 2020), https://www.ajc.com/news/crime--law/georgia-gang-database-has-law-enforcement-hopeful-critics-worried/TJelxnLcBflQQeWpTjY6EJ [https://perma.cc/YQ6B-2ATL], which states that "[c]riteria include admitting gang affiliation; having gang tattoos; displaying gang signs personally or in graffiti; wearing clothing, colors, jewelry and/or bandannas believed to be 'gang dress;' possessing or being referenced in gang documents; being seen with gang members; being identified by evidence online, being identified by a reliable source or being arrested on a gang crime charge or being suspected of a gang crime."

Second, law enforcement already partners with private companies to access automated biometric-identity verification and investigation tools. *See* Kashmir Hill, *The Secretive Company That Might End Privacy as We Know It*, N.Y. TIMES (Mar. 18, 2021), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html [https://perma.cc/MVD9-Y3SH].

Third, it was recently reported that Palantir already sells capabilities very similar to the hypothetical ones described below to law enforcement. Haskins, *supra*.

89. This Feature will explore several scenarios where the data collected about one person may be used against another person. Through these scenarios, we can examine the social effects of many common and widespread data practices. For example, data about one person may affect another person via the linkage of two datasets, by revealing data about social networks or genetic information that are by definition shared information, or by applying a prediction algorithm trained off of the data of one person and used against another. This relational effect can have a considerable outsized impact. For example, Cambridge Analytica directly collected data from the 270,000 people who downloaded the "thisisyourdigitallife" application. Because Cambridge Analytica was able to receive those people's social-network data (the profiles of their friends and family), they obtained the profile information of about eighty-seven million users (70.6 million in the United States). This information was used to train an ad-targeting program that delivered microtargeted political advertisements to some portion of Facebook's 190 million United States users (as well as users in the United Kingdom and elsewhere), based on their likelihood to respond to a given advertisement. *See* Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html [https://perma.cc/UUX4-HGBA]; Hannes Grassegger & Mikael Krogerus, *The Data That Turned the World Upside Down*, VICE (Jan. 28, 2017, 9:15 AM), https://www.vice.com/en/article/mg9vvn/how-our-likes-helped-trump-win [https://perma.cc/9748-

Standard privacy critiques of data flows like this one emphasize not only the significant stakes of this data flow for both Adam and Ben, but also that Adam's data was collected under highly coercive conditions (that is, while Adam was detained in prison) for a purpose (to identify other gang members) with which he may not agree and over which he has no say. Adam's lack of agency at the point of his data's collection is accorded significant moral and legal relevance in critiques of biometric surveillance.[90]

Adam's and Ben's interests are sufficiently aligned such that enhancing Adam's legal rights vis-à-vis TattooView AI would likely also protect Ben's interests. If Adam is granted a robust right to refuse inclusion of his tattoos in TAG-IMAGE, he is likely to exercise that right. Then, his tattoo-image data cannot be used to detain Ben.

But consider an alternative scenario: TattooView AI develops its tattoo recognition algorithm not from the FBI's TAG-IMAGE dataset, but from a dataset it obtained when TattooView AI purchased TattooID. TattooID is a social platform where tattoo enthusiasts can share photos of their tattoos, tag their tattoo artist, and search for designs. Suppose that Adam, a former gang member who regrets his gang involvement, voluntarily shares his tattoo images on TattooID, and tags them as tattoos related to gang affiliation in the hopes that they can help identify other gang members.

In this alternative scenario, individualist conceptions of how this data may harm Adam do not capture the way this data flow affects Ben. Adam was not coerced into sharing this data, but instead did so willingly. Moreover, the purposes to which this data is being applied align with Adam's intent in sharing it and would, in his view, represent a valid outcome. And yet the fact remains that

---

LBQM]; Issie Lapowsky, *Facebook Exposed 87 Million Users to Cambridge Analytica*, WIRED (Apr. 4, 2018, 5:43 PM), https://www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica [https://perma.cc/V9A4-UW6D]; Sam Meredith, *Facebook-Cambridge Analytica: A Timeline of the Data Hijacking Scandal*, CNBC (Apr. 10, 2018, 9:51 AM EDT), https://www.cnbc.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html [https://perma.cc/3BTL-2QLA]; Owen Bowcott & Alex Hern, *Facebook and Cambridge Analytica Face Class Action Lawsuit*, GUARDIAN (Apr. 10, 2018, 11:45 AM EDT), https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit [https://perma.cc/UUN2-FGHE].

90. *See, e.g.*, Complaint at 3, ACLU v. Clearview AI, Inc., No. 9337839 (Cir. Ct. Cook Cnty., May 28, 2020), https://www.aclu.org/legal-document/aclu-v-clearview-ai-complaint [https://perma.cc/3NA9-VJMX] ("The ability to control their biometric identifiers and to move about in public, free from the threat of surreptitious unmasking or surveillance, is essential to Plaintiffs' members, clients, and program participants in Illinois . . . . Clearview has captured more than three billion faceprints from images available online, all without the knowledge—much less the consent—of those pictured.").

Ben faces significant consequences from this data flow. To the extent Ben's interests are of legal, as well as normative relevance, this presents a problem for data-governance law.

B. Mapping Data Social Relations Along Vertical and Horizontal Axes

The relationships that arise among data subjects, data producers, and the third parties impacted by data use can be mapped along two axes.

Along the vertical axis lies the data relation between an individual data subject and an individual data collector (also known as a data processor). The vertical data relation describes the relationship between Adam and TattooView AI, when Adam agrees to the terms of data collection laid out by TattooView AI and shares his data with them. This vertical data relation structures the process whereby data subjects exchange data about themselves for the digital services the data collector provides.

This vertical social relation is expressed both technically via the flow of data from a data subject to a data collector and legally via the contractual terms that structure the terms of exchange between data subject and data collector, as well as the background consumer- and privacy-law regimes that allocate privileges, claims, and duties between the two parties. This vertical data relation is, in some sense, well understood in data-governance law. As will be discussed in greater detail in Part III, proposals for data-governance reform are attentive to how the law governs this vertical relation, how it may structure unequal relations among data subjects and data producers, and how duties and rights between these two parties may be reallocated to address this imbalance.

The horizontal axis describes how data production relates data subjects not to data collectors, but to one another and to others that share relevant population features with the data subject. The relationship between Adam and Ben describes a horizontal data relation.

This horizontal relation is expressed technically through informational infrastructures that make sense of data subjects via group classification and that operationalize classifications to act back on subjects. These technical expressions apprehend (and in apprehending, help to define) the social fact of group identity via shared preferences, social patterns, and behaviors that make people similar to one another. For example, the horizontal data relation between Adam and Ben apprehends a particular social meaning based on their shared tattoo. This horizontal data relation structures a social process whereby a relevant shared feature (that is, a tattoo) is operationalized to make a prediction and define a social meaning (that is, gang membership) and act back on a group member (Ben) according to this grouping.

Horizontal relations are not actually one-to-one relations between data subjects like Adam and those impacted by data flows like Ben. They are population-based relations. For instance, sharing his tattoo image puts Adam in horizontal relation not only with Ben, but also with everyone who has his tattoo and may be acted upon on the basis of this shared feature. The same holds the other way around. Ben is not only in horizontal relation with Adam, but also with everyone who has a relevant population feature in common with him (in this case, his tattoo) and has shared data about this feature with a data collector.[91] To make the discussion below clearer, let's call the group to which Ben belongs "P_{use}," to denote the population of people on which Adam's tattoo image data is used, and refer to the other individuals (the "other Bens") in this population as B_n.[92] We can call the group to which Adam belongs "$P_{collect}$," to denote the population of people from whom tattoo image data is being collected, and refer to other individuals (the "other Adams") in this population as A_n.

These population-level relations give rise to population-level interests along the horizontal relation. For example, we can understand Ben's interest in Adam's data collection as one instance of the more general interest of P_{use} in $P_{collect}$'s data collection. This interest, unlike those along the vertical relation, does not reduce to the individual provenance of the data. Ben's interest in $P_{collect}$'s data sharing is based on the effect that the use of this data will have on him. This use may occur regardless of whether this data was collected from him, from Adam, or from someone else. In this sense, it does not matter *who* the data "came" from, but *what* such data says about Ben, and *how* such meaning is used to act upon Ben. This is the population-level interest Ben (and others like Ben) have in data that apprehends a relevant shared population feature about them. As the example shows, this interest may arise from data Ben shares, data Adam shares, or data someone else shares. Each individual instance of this interest may be weak, but they occur at scale throughout the data-production economy and link individuals to many other individuals via webs of horizontal relation.

One way to understand data governance's unsatisfying response to downstream social effects from data collection, what I call the "sociality problem," is data-governance law's conceptual commitment to individualism (DIM). This commitment focuses the relevant analysis on how data production may harm data subjects and develops legal responses to such harm. While this commitment may result in improvements to the vertical data relation between data subjects like Adam and data collectors like TattooView AI, it does not address the role that horizontal data relations play in producing social value and social risk. This has several significant consequences discussed in greater detail below.

---

91. $P_{collect}$ = [Adam, $A_1, A_2, A_3 \ldots$], where A_n = other individuals in $P_{collect}$.

92. P_{use} = [Ben, $B_1, B_2, B_3 \ldots$], where B_n = other individuals in P_{use}.

C. *The Importance of Horizontal Data Relations in the Digital Economy*

While horizontal data relations are minimally relevant to data-governance law, they are central to how data production produces both social value and social risk. Data production for the digital economy is deeply—even fundamentally—relational.

Data flows are quite literally structured, collected, and produced so as to relate people to one another.[93] Data flows are useful when they relate people to one another. Data flows are designed to represent the ways that people are like one another and reveal meaningful things about one another: how we are alike biologically, interpersonally, politically, and economically.[94]

---

93. *See* GREEN, *supra* note 7, at 93-103; Rob Kitchin, *Big Data, New Epistemologies and Paradigm Shifts*, BIG DATA & SOC'Y, Apr.-June 2014, at 2 (detailing how Big Data is not simply denoted by volume, but is, among other features, denoted by being relational in nature, which here means "containing common fields that enable the conjoining of different data sets," and flexible, "holding the traits of *extensionality* (can add new fields easily) and *scalability* (can expand in size rapidly)"); Danah Boyd & Kate Crawford, *Critical Questions for Big Data*, 15 INFO., COMMC'N & SOC'Y 662, 670-71 (2012). For more on the utility of data because of its ability to reveal information on others, see, for example, Sebastian Benthall & Jake Goldenfein, *Data Science and the Decline of Liberal Law and Ethics* 7-8 (Oct. 2, 2020) (unpublished manuscript), https://ssrn.com/abstract=3632577 [https://perma.cc/C2Q2-8D9L].

94. Almost all data harvested from an individual person or personal device has the capacity to be relational. Social-media data reveals information (such as preferences and observations) not just about an individual, but also about her social networks. This information can have political as well as social consequences. For example, network data can be used to probabilistically identify support or opposition for a political candidate or position to target political advertising or get-out-the-vote efforts. *See* Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle & James H. Fowler, *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, 489 NATURE 295, 295-98 (2012). Network data can be used to predict a credit score and help proactively track and target a disease, suicides, and gun violence. Ben Green, Thibaut Horel & Andrew V. Papachristos, *Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014*, 177 JAMA INTERNAL MED., 326, 326-33 (2017). Genetic data from a consumer genome test reveals information about one's relatives that can help them detect disease early while also placing them in political and legal projects of reconciliation with and reparations for past racial discrimination. *See* ALONDRA NELSON, THE SOCIAL LIFE OF DNA: RACE, REPARATIONS, AND RECONCILIATION AFTER THE GENOME, at xii-xiii (2016). Location data reveals information about one's household. In fact, almost all data harvested from one person that can be used to make a prediction about them or attempt to change their behavior can be utilized, in the form of a behavioral model, to make a prediction or attempt to change the behavior of others. For more on the utility of data because of its ability to reveal information on others, see, for example, Benthall & Goldenfein, *supra* note 93. Institutional-economics literature extols the competitive value of data via tailoring, prediction, personalization, nudging, and marketplace design. *See, e.g.*, MIT TECH. REV. CUSTOM, *supra* note 25, at 2 ("Data is now a form of capital,

Data flows classify and sort people along particular categories of group membership. This process of sorting and classifying is how an individual becomes rendered as a data subject and is how economic value from production is realized. In other words, data about individuals is useful in the digital economy because it helps to define relevant group categories. These categorizations are operationalized to make sense of people on the basis of their classifications and to act back on such insights.

Data about populations is used to develop models to predict and change behavior, to gain intimate consumer or competitor knowledge for market advantage, and to retain greater surplus value.[95] Such activities demonstrate an orientation toward data subjects that recognizes them not as individuals, but as members of groups that are constituted via their shared features and common patterns of behavior.[96] This process recasts people as assemblages of their social relations and group behaviors and apprehends data subjects as patterns of behavior derived from group-based insights. This basic approach is what makes behavioral targeting, prediction tasks, at-scale risk assessment, and modulated feedback systems both possible and profitable.[97]

Data's relationality is central to how data collection produces economic value. This distinguishes the value of Adam's data for the machine-learning (ML) or artificial-intelligence (AI) applications of the contemporary digital economy from the value personal data has for older forms of consumer surveillance (and that inform the law's current approach to data privacy). Prior to the widespread availability of large-dataset computing technology, data about a data subject like Adam may have been valuable because it helped businesses, employers, the government, and insurers know things about Adam. And to some extent, this is still what makes data about Adam valuable.

But what makes data about Adam particularly and distinctly valuable for the contemporary digital economy is its capacity to help companies make predictions or change the behaviors of *others* based on relevant population features they share with Adam — in other words, on the basis of at-scale population-level horizontal

on the same level as financial capital in terms of generating new digital products and services."). This literature aligns with conceptions of human behavior as predictive and probabilistic in cybernetics. *See* JAMES R. BENIGER, THE CONTROL REVOLUTION: TECHNOLOGICAL AND ECONOMIC ORIGINS OF THE INFORMATION SOCIETY (1986). On the concept of data enclosure, see COHEN, *supra* note 3, at 62-63.

95. Salomé Viljoen, Jake Goldenfein & Lee McGuigan, *Design Choices: Mechanism Design and Platform Capitalism*, BIG DATA & SOC'Y, July-Dec. 2021, at 1, 5.

96. *See* JAKE GOLDENFEIN, MONITORING LAWS: PROFILING AND IDENTITY IN THE WORLD STATE 101-02 (2019); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 191-94 (2015); Benthall & Goldenfein, *supra* note 93, at 893.

97. PASQUALE, *supra* note 96.

data relations. It is this relational value of data that drives much of the imperatives to data access, processing, and use. The distinctive feature of ML- and AI-based systems is that they can be used to know things about Adam that Adam does not know, by inferring back to Adam from A_n. And, of greater legal significance (or concern), data from A_n can be used to train models that "know" things about B_n, a population that may not be in any vertical relation with the system's owner. This is the key shift of at-scale data analysis, as compared to prior digital data collection and use approaches that did not have access to the scope and degree of data aggregation, computation capabilities, and inference models that typify digital economic activity in the past decade. It also highlights the importance of horizontal data relations not only for expressing an expanded set of interests in data flows, but also for structuring the incentives of data collectors along vertical data relations with data subjects.

Two implications follow from recognizing the significance of data's relationality in the digital economy. First, conceiving of data's horizontal relationality as incidental to the task of managing data production is wrong. Data's horizontal relationality does result in observable externality effects (from the perspective of the data subject and from that of status quo data governance); however, conceiving of these effects as "external" to the purposes and uses of data that drive entities to transact for it is incorrect.[98] Enacting horizontal relations is not like producing pollution; if polluters could "magic away" pollution they likely would (if only to save themselves some reputational harm). But the same cannot be said for data producers: data's relationality is central to the business of data production and constitutes much of what makes data production economically valuable in the first place.

Second, data's aggregate effects amplify the consequences of this disconnect. In a typical data flow, any one individual's data is essentially meaningless, and the marginal cost of any one individual defecting from collection is very low.[99]

---

98. To be clear, the general descriptive claim that data production results in externalities, that is, that data production produces social costs that are not reflected in the individual transaction, is one this Feature agrees with. *See* Solove, *supra* note 40, at 1880-83; Reidenberg et al., *supra* note 74, at 485-86. However, descriptively noting that an externality exists has limited analytic value in illuminating the structural conditions behind, and inegalitarian features of, data production's social effects.

99. *See* Michael Mandel, *The Economic Impact of Data: Why Data Is Not Like Oil*, PROGRESSIVE POL'Y INST. 6 (2017), https://www.progressivepolicy.org/wp-content/uploads/2017/07/PowerofData-Report_2017.pdf [https://perma.cc/Y8NL-DEED] ("[U]nused data, by itself, has uncertain economic value. Its value depends on how it is combined and used with other data."); *see also* POSNER & WEYL, *supra* note 17, at 205-09 (describing the concept of people as "data producers," who, in aggregate, power the digital economy); Alessandro Acquisti & Jens Grossklags, *Privacy and Rationality in Individual Decision Making*, IEEE SEC. & PRIV., Jan./Feb.

Yet in aggregate, data is highly valuable and grows in value the more data can be combined with other kinds of data.[100] Across many different fields of algorithmic development and machine learning—from computer vision to natural language processing to adversarial machine learning—the rule of thumb is that quality and quantity of data in a model's training set are the biggest determinant of overall performance.[101] More data means better models, which results in digital products that make better predictions about both data subjects as well as others who share relevant features with data subjects. Large-scale data collection and aggregation therefore become key competitive advantages in the digital economy.

Treating data's relationality as an accidental byproduct of data creation misdiagnoses a feature as a bug. The combination of relational and aggregate effects from data production drives companies to collect as much data as possible from data subjects. Data subjects are in turn poorly equipped to exert meaningful coercive force back on to data collectors in the face of collectors' strong incentives to obtain such data. However, the issue is not simply a mismatch in the relative coercive power between parties, but rather the wide range of interests that are not represented in these transactions at all, even while the economic benefits of exploiting these interests motivate the data-collection practices of digital firms.

2005, at 24-30 (explaining how incomplete information, bounded rationality, and systemic psychological deviations from rationality affect individual privacy-sensitive behavior); Alessandro Acquisti, Leslie K. John & George Loewenstein, *What Is Privacy Worth?*, 42 J. LEGAL STUD. 249, 249-74 (2013) [hereinafter Acquisti et al., *What Is Privacy Worth?*] (investigating how individuals conduct privacy valuations); Alessandro Acquisti, Curtis Taylor & Liad Wagman, *The Economics of Privacy*, 54 J. ECON. LITERATURE 442, 442-92 (2016) [hereinafter Acquisti et al., *The Economics of Privacy*] (highlighting how consumers' economic analyses of privacy have evolved over time); Imanol Arrieta Ibarra, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier & E. Glen Weyl, *Should We Treat Data as Labor? Moving Beyond "Free,"* 108 AM. ECON. ASS'N PAPERS & PROC. 38, 38-42 (2017) (arguing for a paradigm where data is not seen as capital, but as labor).

100. For a detailed discussion of how data creates value, see *supra* Section I.A. In the machine-learning community, it is commonly understood that as datasets grow larger and more readily aggregated with other sources of data, they grow more valuable. *See* sources cited *supra* note 94; POSNER & WEYL, *supra* note 17.

101. *See* Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, 55 COMMC'NS ACM 78, 84 (2012) ("More data beats a cleverer algorithm."); *see also* Alon Halevy, Peter Norvig & Fernando Pereira, *The Unreasonable Effectiveness of Data*, 24 IEEE INTELLIGENT SYS. 8, 8-12 (2009) ("A trillion-word corpus . . . captures even very rare aspects of human behavior."); Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton & Alex Hanna, *Data and Its (Dis)contents: A Survey of Dataset Development and Use in Machine Learning Research* 1 (NeurIPS 2020 Workshop: ML Retrospectives, Surveys & Meta-Analyses, Working Paper, 2020) ("The importance of datasets for machine learning research cannot be overstated."); Chen Sun, Abhinav Shrivastava, Saurabh Singh & Abhinav Gupta, *Revisiting Unreasonable Effectiveness of Data in Deep Learning Era*, *in* PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION 843, 843-52 (2017) ("The success of deep learning . . . can be attributed to . . . availability of large-scale labeled data.").

The prevalence of horizontal interests in data thus creates a structural mismatch in vertical relations between data subjects and data collectors. Data subjects possess only a fraction of the interests in a given data flow—and, as described above, many of their interests in information do not reduce to their vertical transaction with a data collector either. Meanwhile, data collectors are highly motivated to collect as much data from as many data subjects as possible in order to realize the considerable benefits that accrue from exploiting the insights of horizontal data relations. Without attending to horizontal relations in data-governance law, the interests they represent and the behaviors they motivate from data collectors cannot be fully accounted. Misdiagnosing these effects as incidental to the task of preventing further privacy erosion risks developing reforms that are not up to the task of disciplining excessive or overly risky data production.

D. The Absence of Horizontal Data Relations in Data-Governance Law

While horizontal data relations are of primary importance in explaining why data collectors develop infrastructures to collect and monetize data flows, they do not feature much, if at all, in how current data-governance law allocates claims, privileges, and duties among actors in the digital economy. Many of the relevant interests in data production that accrue along these population-level relations are unrepresented in data-governance law.

This has both practical and normative implications. First, as a practical matter, the absence of legal interests for horizontal data relations leaves the law out of step with the importance of these relations for the digital economy. As discussed above, this may preclude effective regulation of vertical relations as well. The imperatives to relate individuals along the horizontal axis motivate data collectors and influence the conditions of exchange between them and data subjects. Horizontal relations are therefore relevant to the task of regulating subject-data collector vertical relations.

Second, the absence of horizontal data relations in law may cause data-governance law to miss—or misconceive—how data production results in particular kinds of injustice. Because population-level interests are not represented, data-governance law is not indexing forms of injustice that operate via horizonal relations. This misconception may also lead to regimes of data governance that inadvertently foreclose socially beneficial forms of data production. This second implication is discussed in greater detail below.

The legal marginality of horizonal data relations leaves many consequences of data production unaccounted for in data-governance law. This includes externalities (such as Ben's lack of representation) in how the law accounts the sum of risks and benefits in the data flow from Adam to TattooView AI. But it also

leaves unaddressed distributive effects: how data flows spread the benefits and risks of data production unevenly among actors in the digital economy, often along the lines of group identities that serve to inscribe forms of oppression and domination. For instance, if Ben is Black, the incapacity of data-governance law to represent Ben's interests in Adam's data flow presents problems that are of a different (arguably more significant) normative quality, given the way this data flow materializes a racialized social process.

Certain forms of data production may equally subject individuals to coercive forms of data collection, but lead to unequally harsh consequences from the resulting data flows. While coercive collection practices may generally constitute unjust vertical relations, the resulting horizonal relations may enact normatively distinct group-based forms of oppression. For example, consider the recent Immigration and Customs Enforcement (ICE) and Customs and Border Protection (CBP) purchase of mobile-location data from the company Venntel to identify and arrest suspected undocumented immigrants on the basis of mobile-phone activity in remote borderlands.[102] Location data in Venntel's database tracks location information from millions of mobile phones, and is drawn from mobile applications like games and weather apps that request access to users' location data. ICE has also purchased licenses from Clearview AI, a facial-recognition company that recently drew public scrutiny for its widespread use among law-enforcement agencies and dubious — possibly even illegal — data-collection practices.[103] In both instances, millions of data subjects are subject to data-collection practices by Clearview and Venntel that may fail to meet the standard of meaningful consent. Many data subjects may find these data practices unfair or unjust, and express interest in reforming data-collection practices to address them.[104]

---

102. *See* Byron Tao & Michelle Hackman, *Federal Agencies Use Cellphone Location Data for Immigration Enforcement*, WALL ST. J. (Feb. 7, 2020), https://www.wsj.com/articles/federal-agencies-use-cellphone-location-data-for-immigration-enforcement-11581078600?mod=hp_lead_pos5 [https://perma.cc/RTS5-2DRT]; Blest, *supra* note 7.

103. *See* Kim Lyons, *ICE Just Signed a Contract with Facial Recognition Company Clearview AI*, VERGE (Aug. 14, 2020, 3:19 PM ET), https://www.theverge.com/2020/8/14/21368930/clearview-ai-ice-contract-privacy-immigration [https://perma.cc/JMY3-NB35]. Clearview AI built its facial-recognition database by scraping publicly available face images from the web, in violation of Illinois's Biometric Information Privacy Act, which requires companies to obtain notice from consumers before collecting and using their biometric information. Biometric Information Privacy Act (BIPA), 2008 Ill. Legis. Serv. P.A. 095-994 (codified as amended at 740 ILL. COMP. STAT. 14/15 (2021)).

104. Andrew Perrin, *Half of Americans Have Decided Not to Use a Product or Service Because of Privacy Concerns*, PEW RSCH. CTR. (Apr. 14, 2020), https://www.pewresearch.org/fact-tank/2020/04/14/half-of-americans-have-decided-not-to-use-a-product-or-service-because-of-privacy-concerns [https://perma.cc/6LXC-B6JH]; Bruce Schneier, *We're Banning Facial Recognition. We're Missing the Point*, N.Y. TIMES (Jan. 20, 2020), https://www.nytimes.com/2020/01/20/opinion/facial-recognition-ban-privacy.html [https://perma.cc/K3VP-39K8].

However, the risks associated with how this data is used fall unevenly among the population of those they impact. This in turn presents a different class of harm than simply unjust conditions of data collection. The Venntel data flow enacts horizontal data relations whereby a relevant shared feature (i.e., movement patterns) is operationalized to make a prediction (undocumented immigrant) and act back on a group member according to this categorization (detaining them). This amplifies the stakes of this data flow, as well as the shared feature it acts upon, on the basis of membership in a socially oppressed group. Individuals who, due to their race, ethnicity, religion, or language, are subject to heightened scrutiny from immigration officials face disproportionate risks to themselves and others like them from having their movement patterns — or those of people like them — apprehended via these data flows.

This is not due to some inherently oppressive feature of movement patterns. Instead "movement patterns" as a relevant shared-population feature become constitutive of how members of this population are socially defined and acted upon in oppressive ways. Movement patterns become a useful identifying feature for undocumented immigrants and are then used to detain group members on the basis of their immigration status. In other words, this horizontal relation materializes a social process of oppression. If one is *not* a member of the relevant group (undocumented immigrant), one faces negligible risk of this kind of social informational harm, even if one's location data is being collected.

These unevenly distributed risks suggest that even where data subjects are subject to equal conditions of collection, the benefits and risks from use may be spread unevenly, amplifying the harmful social consequences of minority group memberships. This harm is normatively distinct from potentially unjust data collection. It locates injustice in the social process this data flow enacts, not the conditions under which it was collected. Reducing concerns over this data flow to the (unjust) conditions of collection alone thus underrepresents both the overall stakes of collection, and the normative significance of how and why such stakes are distributed unevenly.

More socially advantaged groups may even engage in voluntary data collection that benefits them, while harming socially disadvantaged groups. The horizontal relations between voluntary data subjects and involuntary third parties may materialize social processes that amplify the (oppressive) differences between groups. For example, consider a scenario where a homeowner (let's call her Alice) voluntarily installs the Amazon Ring, a popular internet- and video-enabled doorbell that allows residents to remotely record their front porch and speak to individuals.[105] Like many Ring users, Alice also joins Ring's Neighbors

105. *Ring*, AMAZON, https://www.amazon.com/stores/Ring/Ring/page/77B53039-540E-4816-BABB-49AA21285FCF [https://perma.cc/BXC5-2RKJ].

app, which allows her to receive and post real-time crime and safety alerts.[106] Alice knows and approves of the partnership between Neighbors and her local law-enforcement agency.

Alice has two neighbors, Beatrice (who is white) and Cara (who is Black). Because data collected from Alice's Ring may be used to report and act on Beatrice and Cara on the basis of a shared feature (i.e., they all live in the same small radius in which Ring-based alerts may lead to intervention), both are in horizontal data relationships with Alice. Both Beatrice and Cara are third parties to Alice's transaction with Ring. Both bear externalities from Alice's relationship with Ring due to their unrepresented interests in this data flow. Both may benefit from having their porches under Alice's surveillance. And both also incur some risk: shared population data about them flows from Alice's Ring to the Neighbor App, Amazon, and local law enforcement. But Cara faces greater risk of possible violence from this horizontal data relation than does Beatrice. Her data relation with Alice is one way the preexisting unjust social processes of racial hierarchy are materialized. These materialized social processes are what make it more likely that Alice's surveillance leads to violence against Cara from law enforcement or other neighbors. The two horizontal relations between Alice and Beatrice and Alice and Cara thus carry normatively distinct meanings: one may result in the productive or distributive inefficiencies that arise due to externalities, while the other may serve to reproduce or amplify racism.

These disproportionate risks suggest that even when data subjects voluntarily consent to data collection, relevant horizontal relations remain unrepresented in law in ways that can amplify the harmful and subordinating consequences of marginal group membership.

---

106. RING, https://ring.com/neighbors [https://perma.cc/4CXY-6M9N]. Neighbors have entered into video-sharing partnerships with over 1,300 local law-enforcement agencies. *See Atlas of Surveillance*, ELEC. FRONTIER FOUND., https://atlasofsurveillance.org [https: //perma.cc/G2VL-2DCQ]; *see also* Khaleda Rahman, *Police Are Monitoring Black Lives Matter Protests with Ring Doorbell Data and Drones, Activists Say*, NEWSWEEK (Aug. 9, 2020, 10:46 AM EDT), https://www.newsweek.com/amazon-ring-drones-monitor-protests-1523856 [https://perma.cc/KK6U-N4AK] (noting these agreements grant departments "special access" to Amazon Ring's Neighbors app and its crime and safety alerts, drawing out the implications of this access during the racial-justice protests of 2020, as well as noting that the vertical relation between owners and Amazon Ring may not be sufficient to circumvent police obtaining footage); Rani Molla, *How Amazon's Ring Is Creating a Surveillance Network with Video Doorbells*, VOX (Jan. 28, 2020, 12:08 PM EST), https://www.vox.com/2019/9/5/20849846/amazon-ring-explainer-video-doorbell-hacks [https://perma.cc/7ASK-FN8T].

III. DIM REFORMS AND THEIR CONCEPTUAL LIMITS

This Part evaluates two prominent legal-reform proposals that have emerged in response to concerns over datafication: propertarian proposals and dignitarian proposals. Propertarian proposals respond to growing wealth inequality in the data economy by formalizing individual propertarian rights over data as a personal asset. Dignitarian proposals respond to excessive data extraction's threat to individual autonomy by granting fundamental-rights protections to data as an extension of personal selfhood. While both types of reforms have some merit, they both suffer from a common conceptual flaw. Both attempt to reduce legal interests in information to individualist claims subject to individualist remedies that are structurally incapable of representing the horizontal, population-level interests of data production. This in turn allows significant forms of social informational harm to go unaddressed and may foreclose socially valuable forms of data production.

A. Propertarian Data-Governance Reform

In response to the harms of data extraction, scholars, activists, technologists and even presidential candidates have all advanced proposals for data-governance reform. Many of these reformers are motivated by the connection between data extraction and wealth accumulation—and hope to redistribute wealth more broadly among data subjects and data processors.

Sir Tim Berners-Lee (inventor of the World Wide Web), for example, created Solid, a web-decentralization project, out of concern for how data extraction fuels a growing power imbalance online.[107] He notes that "for all the good we've achieved, the web has evolved into an engine of inequity and division; swayed by powerful forces who use it for their own agendas."[108] Solid "aims to radically change the way Web applications work today, resulting in true data ownership as well as improved privacy."[109] Solid is especially popular within the blockchain community's #ownyourdata movement.

Glen Weyl (an economist) and Eric Posner (a legal scholar) have similarly introduced a proposal called Radical Markets, which seeks to introduce a labor

---

107. Solid aims to respond to the de facto enclosure of data via a system that ensures personal-data control via local storage, mediated by a series of contractual agreements for access to the user's data. *See Solid*, INRUPT, https://inrupt.com/solid [https://perma.cc/LR5F-DFTV].

108. Tim Berners-Lee, *One Small Step for the Web*, MEDIUM (Sept. 29, 2018), https://medium.com/@timberners_lee/one-small-step-for-the-web-87f92217d085 [https://perma.cc/H5QK-WT5T].

109. *What Is Solid?*, SOLID PROJECT MIT, https://solid.mit.edu [https://perma.cc/S342-7774].

market for data.[110] Weyl and others advocate for data-as-labor as a response to inequality. In so doing, they aim to disrupt the digital economy's "technofeudalism," where the uncompensated fruit of data laborers is "distributed to a small number of wealthy savants rather than to the masses."[111]

Progressive politicians, concerned over inequality in the information economy, have also advanced similar proposals. Former presidential candidate Andrew Yang, for example, included a right to data property in his campaign platform.[112] And he has recently launched the Data Dividend Project to push companies like Facebook and Google to pay users a "data dividend" for the wealth their data capital generates.[113] Representative Alexandria Ocasio-Cortez has also posited data ownership as a solution to inequality, tweeting: "[T]he reason many tech platforms have created billionaires is [because] they track you without your knowledge, amass your personal data[]& sell it without your express consent. You don't own your data, & you should."[114]

Their proposals all advance a version of data-governance reform that grants a propertarian entitlement to data. Propertarian reforms formalize the right to data as an individual's entitlement to their data-assets. Most reforms propose a property right over data about the subject, in which the data subject may then sell usage or full ownership rights. Alternatively, data production may be conceived of as a form of the subject's labor that entitles the data subject to command a wage in a data-labor market.

Propertarian data reform posits a particular legal solution to the problems of data extraction that transforms data *about* the subject into an asset that generates

---

110. Jaron Lanier was one of the earliest to propose a data-as-labor conception. He similarly "worries about the distributional and social consequences of the failure to pay for data and online creative production." POSNER & WEYL, *supra* note 17, at 222; *see* JARON LANIER, WHO OWNS THE FUTURE (2013). Lanier's proposal is taken up by Glen Weyl, Eric Posner, and others as preferable over data as property or capital, because it captures the role individuals have in generating value in the data economy. On this view, it is necessary to conceive of data as labor, not capital, to restore a functioning market for user contributions. POSNER & WEYL, *supra* note 17, at 209; *see* Imanol Arrieta Ibarra, Leonard Goff, Diego Jiménez Hernández, Jaron Lanier & E. Glen Weyl, *Should We Treat Data as Labor? Let's Open Up the Discussion*, BROOKINGS INST. (Feb. 21, 2018), https://www.brookings.edu/blog/techtank/2018/02/21/should-we-treat-data-as-labor-lets-open-up-the-discussion [https://perma.cc/KLR5-MEV2].

111. POSNER & WEYL, *supra* note 17, at 231, 209; *see also* Arrieta Ibarra et al., *supra* note 99, at 38 (noting that data as free data "skews distribution of financial returns from the data economy").

112. *See Data as a Property Right*, YANG2020, https://www.yang2020.com/policies/data-property-right [https://perma.cc/RL3Q-J4GC].

113. DATA DIVIDEND PROJECT, https://datadividendproject.com/aboutus [https://perma.cc/6HCN-5PTP].

114. Alexandria Ocasio-Cortez (@AOC), TWITTER (Feb. 19, 2020, 11:43 PM), https://twitter.com/AOC/status/1230352135335940096 [https://perma.cc/E2KX-GGX9].

wealth *for* the subject.[115] On this view, data is already being "coded" as quasi capital in law (through a combination of contractual agreements and trade-secrecy law) in a manner that serves to create wealth for its holders, but excludes the individuals from whom data originated.[116] The problem is not with the conceptualization of data as capital *per se*, but rather with who has legal rights to benefit from that capital. As a legal matter, enacting propertarian reforms would code data with features considered more amenable to wealth creation for data subjects.[117] Data governance, therefore, becomes the governance (via contract law, property law, employment law, and labor law) of property relations or wage relations. This translates into a legal reform agenda to change the legal code being *applied* to data assets, not to reject the concept of data *as* an asset.

Moving from de facto to de jure property rights over data is meant to secure several benefits classically associated with propertarian reforms. First, it clarifies rights of self-determination and control over data by allocating legal entitlements over data to data-subjects. I call this the "data control claim," which has the corollary effect of establishing at least some alienable claims to data. Second, it allows for bargaining between data subjects and collectors in a marketplace for personal data with the aim of achieving a Pareto-efficient allocation of the benefits (and hence, Pareto-efficient levels of production and consumption) of data extraction. I call this the "market efficiency claim." Third, by compensating individuals for the value they help to create, such entitlements are meant to spread

---

115. *See, e.g.*, POSNER & WEYL, *supra* note 17, at 205-07 (dramatizing a scenario in which Facebook pays users for interpersonal information they share with the platform).

116. *See* COHEN, *supra* note 3, at 63. For a detailed treatment of how assets are coded in law to become capital, see KATHARINA PISTOR, THE CODE OF CAPITAL 2-3 (2019) ("Fundamentally, capital is made from two ingredients: an asset, and the legal code . . . . With the right legal coding, any of these assets can be turned into capital and thereby increase its propensity to create wealth for its holder(s).").

117. As Pistor aptly describes, once data is conceived of as an asset of any kind in law, any conceptual distinction between capital (*K*) and labor (*L*) is reduced. PISTOR, *supra* note 116, at 11. In law, both render data as the subject of an exchange relation between data subject and data processor for data's alienable value. As a legal conceptual matter, *L* is easily turned into *K* with a bit of legal engineering. Take, for example, partners in a limited liability partnership (LLP). They contribute their labor to the corporate entity as in-kind services and take out dividends as shareholders in lieu of a salary, thus benefitting from the better legal protections and lower tax rate afforded *K* for the same exact work that would be performed were it coded as *L* instead. *See id*. at 11, 48. Beyond law, the concept of "human capital" in corporate finance, organizational sociology, and labor economics also serves to collapse the conceptual distinction between the contributions of capital assets and labor power to production. *See, e.g.*, Samuel Bowles & Herbert Gintis, *The Problem with Human Capital Theory—A Marxian Critique*, 65 AM. ECON. ASS'N. 74, 74-75 (1975) (discussing how human capital theory allows for fundamental insights regarding labor while also absorbing the explanatory category of labor into a concept of capital); Claudia Goldin, *Human Capital*, *in* HANDBOOK OF CLIOMETRICS 55 (Claude Diebolt & Michael Haupert eds., 2016).

the benefits of the digital economy more widely.[118] I call this the "redistribution claim."

Together, these three claims make propertarian reform an intuitively appealing response to the quasi enclosure and de facto ownership of data resources by technology companies. By formalizing the informal propertarian status of data, such reforms directly counteract the quasi-propertarian claims to personal data flows of large data collectors like Google, Facebook, and Amazon, and directly invalidate the current practice of capturing data value from subjects without compensation.[119]

Propertarian reforms also dovetail nicely with certain diagnoses of—and responses to—problems of competition and fairness in the data political economy.[120] Such views identify either too much corporate control over data assets, or too much concentration in the corporate control of data assets, as key barriers to competition.[121] One popular response to such problems is "data portability," a set of technical interoperability requirements and legal rights that allow users to transfer their data.[122] Under this view, empowering users to "shop" for new

---

118. Propertarian reforms have long been motivated by claims that they can extend material benefits to those who are currently excluded from enjoying them. Development economist Hernando de Soto was a prominent proponent of granting the poor in developing countries property rights as a way to achieve economic security. *See* HERNANDO DE SOTO, THE MYSTERY OF CAPITAL: WHY CAPITALISM TRIUMPHS IN THE WEST AND FAILS EVERYWHERE ELSE (2000). Such rights, he argues, can turn "dead assets" into "live capital," granting owners the opportunity to mortgage land or other assets to invest in new ventures and begin to accrue wealth. *Id.* at 50. This general theory of widespread and shared wealth creation via property rights experienced a "surge" in the 1980s, when the idea of "clear property rights and credible contract enforcement" to create "conditions by which everyone would prosper" was widely adopted by development economists and politicians throughout the world. PISTOR, *supra* note 116, at 1. For discussion of de Soto and the popularity of this reform in development economics, see PISTOR, *supra* note 116, at 1-2, 14.

119. Large platform companies assert quasi-propertarian claims to data flows via their "de facto appropriation and enclosure" of personal data flows. COHEN, *supra* note 3, at 25.

120. *See* Lina Khan, Note, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710 (2017) (examining anticompetitive conduct by digital platforms). The FTC is holding hearings and workshops on the concept. *See FTC Announces September 22 Workshop on Data Portability*, FED. TRADE COMM'N (Mar. 31, 2020), https://www.ftc.gov/news-events/press-releases/2020/03/ftc-announces-september-22-workshop-data-portability [https://perma.cc/3PKP-RE23]. Senators Mark Warner, Josh Hawley, and Richard Blumenthal have introduced a bill to encourage market competition among social-media platforms that includes data portability. Augmenting Compatibility and Competition by Enabling Service Switching (ACCESS) Act of 2019, S. 2658, 116th Cong. (2019) (reintroduced in 2021).

121. *See, e.g.*, STUCKE & GRUNES, *supra* note 17, at 145 ("[T]he concentration of data post-merger could raise an anti-trust concern.").

122. *E.g.*, Gabriel Nicholas, *Taking It with You: Platform Barriers to Entry and the Limits of Data Portability*, 27 MICH. TECH. L. REV. 263 (2021).

digital services would encourage market discipline (due to enhanced user exit options) and give new market entrants the opportunity to attract users and their valuable user data.[123] Data portability combines elements of the data control claim and the market efficiency claim to enhance competitive opportunity via individuals' market actions.

Finally, propertarian reforms respond to an important claim of injustice levied against the digital economy: that individuals play a role in generating a materially valuable resource from which they see no value and which sometimes places them at risk. Calls for entitlement reform, like those discussed above, are often made in response to frustration over the wealth amassed by companies that harvest data for which they pay nothing. At a time when technology companies are widely accused of wielding too much economic and political power over our daily lives, the redistributive claim may contribute to the enduring and widespread appeal of propertarian reforms.[124] Even for those who may view the redistributive claim as purely an instrumental effect of achieving data control and market efficiency, it serves a justificatory role in advocating for propertarian entitlements.

Still, there are several reasons to be skeptical of propertarian data reforms. One is impracticability. Operationalizing the kind of complex and comprehensive micropayments system at the scale required may not be feasible or cost-effective.[125] Moreover, it is empirically unclear whether propertarian solutions would materially address data extraction given current conditions of datafication.[126]

---

123. *Id*. at 276-77. Portability can be seen as an "exit"-enhancing market response. *See* ALBERT O. HIRSCHMAN, EXIT, VOICE, AND LOYALTY: RESPONSES TO DECLINE IN FIRMS, ORGANIZATIONS, AND STATES 52 (1970) (discussing the exit options for durable versus nondurable goods).

124. *See supra* Part I.

125. Critics should be wary of overreliance on convenient arguments of implementation. While such a payments system may appear impracticable from a consumer perspective, as a technical matter, such a system may not be all that different from the highly complex algorithmic-auction platform systems and exchanges through which advertisers purchase views, impressions, and clicks from consumers, which pose similar challenges of managing billions of instantaneous pricing and transaction actions at scale. *See, e.g.*, Viljoen et al., *supra* note 95, at 30.

126. For a more detailed discussion of the limits of data as labor, see Zoë Hitzig, Lily Hu & Salomé Viljoen, *The Technological Politics of Mechanism Design*, 87 U. CHI. L. REV. 95, 101-05 (2019). In short, the mere granting of a labor or property right does not guarantee that the conditions underlying the sale of that labor or property will be non-extractive and uncoerced. The conditions of the current data market do not inspire confidence. Large data collectors are highly concentrated and able to leverage their existing superior knowledge to design exchanges and prices to their advantage. In contrast, data subjects are widely dispersed and isolated from one another, and they have little insight into how data value is created from which to bargain.

Propertarian data reforms may also be unlikely to address the privacy erosion that motivates many concerns over data extraction.[127] Payment provides an additional incentive for people to share data about themselves and thus may further degrade privacy—not only for themselves, but for others as well. A data subject may decide that the risk of their privacy loss is worth the payment provided and thus sell their data in what might appear to be a mutually beneficial exchange. Putting aside the effects this sale has on others, privacy risk is notoriously easy to undervalue at the point of exchange.[128] Privacy risk associated with data is neither static nor linear. It accumulates and grows over time based on the composition effects from multiple sources of data, varied downstream uses, and new

Personal data from any one data subject is essentially valueless, reducing the capacity for individual data subjects to meaningfully exert bargaining power. Moreover, data subjects do not (yet) identify as a common social group from which to build political bargaining power. Finally, datafication does not result in the kind of visceral oppression that may motivate moral outrage and build countervailing power. In contrast with oppressive workplace domination or highly impoverished conditions of production, data extraction is designed to occur as seamlessly and painlessly as possible, transmitting flows of data in parallel with data subjects living their online and offline lives. On the challenges of the U.S. labor market in general, see Matthew Desmond, *Americans Want to Believe Jobs Are the Solution to Poverty. They're Not.*, N.Y. TIMES MAG. (Sept. 11, 2018), https://www.nytimes.com/2018/09/11/magazine/americans-jobs-poverty-homeless.html [https://perma.cc/V64B-R36B].

127. *See, e.g.*, Pamela Samuelson, *Privacy as Intellectual Property?*, 52 STAN. L. REV. 1125, 1136-46 (2000); *see also* Jessica Litman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283, 1295-1301 (2000) (offering a critique of property approaches to privacy); Julie E. Cohen, *Examined Lives: Information Privacy and the Subject as Object*, 52 STAN. L. REV. 1373, 1377-1402 (2000) (critiquing arguments from property rooted in universal concepts of liberty and efficiency); Jane Bambauer, *The Perils of Privacy as Property: The Likely Impact of the GDPR and the CCPA on Innovation and Consumer Welfare*, PROGRAM ON ECON. & PRIV. ANTONIN SCALIA L. SCH. 6-7 (Mar. 12, 2019), https://www.judiciary.senate.gov/imo/media/doc/Bambauer%20Testimony.pdf [https://perma.cc/T7GP-CTWG] (arguing that property-based solutions are unlikely to benefit consumers).

128. *See* Acquisti & Grossklags, *supra* note 99, at 29-32; Acquisti et al., *The Economics of Privacy*, *supra* note 99, at 446-48; Acquisti et al., *What is Privacy Worth?*, *supra* note 99, at 251-52; *cf.* Daniel J. Solove, *The Myth of the Privacy Paradox*, 89 GEO. WASH. L. REV 1, 3 (2021) (responding to the "privacy paradox," a phenomenon observed in behavioral studies where individuals state that privacy matters a great deal to them, but in practice relinquish personal data for very little value). Solove argues that the phenomenon is a paradox only because the underlying studies embrace a faulty logic: believing that how an individual evaluates the risks of a *specific* data exchange in a defined instance and context can be generalized to how such an individual values privacy *generally*. *Id.* at 4. As he notes, "A person does not surrender all privacy when sharing data with others. Many privacy protections remain in place." *Id.* This argument aims to explain the privacy paradox—in other words, why it is that people are observed to undervalue their privacy at the point of exchange. While this argument is not precisely offering an account of the value of data itself, it supports the idea that the observed valuation of a particular data exchange is pervasively undervalued because people do not factor in the value of the background privacy protections that "remain in place," and they do not account for the pluralistic value of privacy for individuals who do not reduce their observed exchange. *Id.*

applications.[129] People tagging online photos of themselves and their friends in 2009, for example, could not have known that companies contracting with law enforcement in 2019 would use that information for facial-recognition products.[130]

Finally, propertarian reforms place greater marginal pressure to sell data on those least able to forego the income it offers, transforming privacy into an even greater privilege than it is today.[131]

Whether in the name of privacy or not, propertarian reforms concede existing processes of data commodification in the digital economy: this ship having sailed, what data subjects can and should secure is their fair share of the value such processes produce.

B. Dignitarian Data Governance

Refusal to concede data commodification lies at the heart of dignitarian critiques of both the status quo and propertarian alternatives. Where propertarian reforms conceive of data as the subject of individual ownership (data as object-like), dignitarian data governance conceives of data as an expression or extension of individual selfhood (data as person-like).[132]

---

129. Aaron Fluitt, Aloni Cohen, Micah Altman, Kobbi Nissim, Salomé Viljoen & Alexandra Wood, *Data Protection's Composition Problem*, 3 EUR. DATA PROT. L. REV. 285, 292 (2019).

130. Lyons, *supra* note 103.

131. Michelle Gilman & Rebecca Green, *The Surveillance Gap: The Harms of Extreme Privacy and Data Marginalization*, 42 N.Y.U. REV. L. & SOC. CHANGE 253, 255 (2018). Proponents of propertarian reforms (to the extent they advocate on the basis of privacy at all) adopt the contested position of privacy as control. Under this view, clarifying data subjects' legal rights over data grants them more control over such data, and is by extension more privacy protective. *See, e.g.*, INRUPT, *supra* note 107 ("Users control which entities and apps can access their data."). For accounts of privacy that contest theories of privacy as control, see NISSENBAUM, *supra* note 6, at 2-3; and Mark Verstraete, *Inseparable Uses*, 99 N.C. L. REV. 427, 429-31 (2021). Verstraete provides an interesting account of privacy as control via a theory of separability that severs the claim of control from a basis in alienability. Verstraete, *supra*, at 431.

132. The European Union's data-governance regime derives its theory of privacy and data protection from Kantian dignitary conceptions of data as an expression of the self, subject to deontological requirements of human dignity. This normative and conceptual account anchors the robust European regime, including its suite of inalienable rights over personal data. *See* Regulation 2016/679, art. 88, of the European Parliament and of the Council of 27 April 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, 84 [hereinafter GDPR], https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [https://perma.cc/2RZ3-KZKT]; Luciano Floridi, *On Human Dignity as a Foundation for the Right to Privacy*, 29 PHIL. & TECH. 307, 307-08 (2016).

Some of the most vivid normative critiques of informational capitalism and privacy erosion invoke dignitarian arguments against datafication. For instance, a highly criticized aspect of information capitalism is that it rewards economic imperatives to apprehend and act on individuals in machine-readable form, often in ways that occur without meaningful consent and for purposes that may violate the wishes of data subjects. Datafication, and the seamless and continual data extraction it relies on, reconstitute individuals into "data doubles," representing them in algorithmically legible forms.[133] In doing so, datafication renders individuals as patterns of behavior, identified as amalgams of categories or classifications (e.g., "Woman," "Millennial," "Lawyer"). Datafication thus violates basic notions of individuals as autonomous beings.

A closely related subject of critique is the affordances of datafication for algorithmic governmentality: the cycle of rendering individuals as patterns of behavior based on certain categories and features, and then algorithmically and iteratively acting on individuals on the basis of these classifications in a state of constant feedback and fine-tuning. This cycle reinscribes algorithmic ways of understanding the subject back onto the subject herself, undermining her capacity for self-formation and the enactment of her self-will.[134]

In response to such concerns, dignitarians like Zuboff argue that datafication and data extraction represent the end of the relationship we enjoy with our innermost selves.[135] The "dark continent" of inner life is invaded and transformed into a "collectivist vision that claims the totality of society."[136] Zuboff diagnoses the injustice of informational capitalism as its endeavor to commodify, colonize,

---

133. COHEN, *supra* note 3, at 67. This Feature's use of the terms "legible" and "legibility" is informed particularly by JAMES C. SCOTT, SEEING LIKE A STATE: HOW CERTAIN SCHEMES TO IMPROVE THE HUMAN CONDITION HAVE FAILED (1998); and MICHEL FOUCAULT, DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON (1977).

134. For an excellent treatment of the subject of how data science theorizes its subject as patterns of behavior, and the disconnect this produces from the subject theorized by law, see Benthall & Goldenfein, *supra* note 93. *See also* Marion Fourcade & Kieran Healy, *Seeing like a Market*, 15 SOCIO-ECON. REV. 9, 10 (2017) (discussing the consequences "of big data-based valuation (of individuals) and value extraction (from individuals) for social stratification"); Dan L. Burk, *Algorithmic Legal Metrics*, 96 NOTRE DAME L. REV. 1147, 1152-53 (2021) (explaining how legal determinations are distorted by algorithmic metrics); Cohen, *supra* note 6, at 1905 (arguing that privacy "protects the situated practices of boundary management through which the capacity for self-determination develops").

135. ZUBOFF, *supra* note 11, at 293, 521; *see also* Zuboff, *supra* note 33, at 76 (describing the "logic and implications of surveillance capitalism as well as 'big data's' foundational role in this new regime"). For an excellent review of Zuboff's enlightenment ideals and their limitations, see Quinn Slobodian, *The False Promise of Enlightenment*, BOS. REV. (May 29, 2019), http://bostonreview.net/class-inequality/quinn-slobodian-false-promise-enlightenment [https://perma.cc/8V43-3V2X].

136. ZUBOFF, *supra* note 11, at 519.

and rule this inner self for profit via monetization schemes that rely on behavioral prediction and control. This new capitalist imperative violates human dignity and destroys personal agency.

Zuboff's repeated invocation of apprehension as violation and behavioral modification as colonization suggests her concern is with datafication itself, not merely with the ends to which it is put or the relations under which it occurs. This account posits datafication as a legibility harm that inflicts on individuals a depth of representation that violates the dignity of their personhood.[137]

This diagnosis is consistent with the dignitarian account of what makes datafication a primary injustice of informational capitalism: that it describes a process of commodification and alienation of the inner self. On this view, rendering a person legible via datafication represents a form of personal violation. Datafication, data extraction, and algorithmic governmentality are wrong because they manipulate people, invade and violate the sanctity of their inner beings, and undermine their capacity to express and enact their free will.[138] To dignitarians, these injustices present ontological and existential threats to personhood and are therefore wrong in their own right, regardless of how the resulting data may be used.

In response to these concerns, dignitarians aim to invigorate legal protections of individual autonomy (or, as is common in the European context, individual dignity) in the digital economy. The strongest such accounts advance legal rights over personal data as akin to natural rights, and thus advocate for fundamental rights to data as an extension of the data subject's moral right to dignity and self-determination.[139] Such rights are contained within the European Union's data-governance regime, which (alongside other affirmative data-

---

137. *Id.* at 521 ("What is at stake here is the human expectation of sovereignty over one's own life and authorship of one's own experience. What is at stake is the inward experience from which we form the will to will and the public spaces to act on that will.").

138. Several other critiques of the digital economy similarly focus on how existing processes of data production undermine individual autonomy. *See, e.g.*, BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY 34 (2018); HARTZOG, *supra* note 61; Susser et al., *supra* note 50; Becky Chao, Eric Null, Brandi Collins-Dexter & Claire Park, *Centering Civil Rights in the Privacy Debate*, OPEN TECH. INST. (Aug. 14, 2019), https://www.newamerica.org/oti/reports/centering-civil-rights-privacy-debate/privacy-is-a-civil-right [https://perma.cc/PDM3-PV3V]; Alvaro Bedoya, *Privacy as Civil Right,* 50 N.M. L. REV. 301, 306 (2020) ("But at its heart, privacy is about human dignity: Whether the government feels it can invade your dignity, and whether the government feels it has to protect the most sensitive, most intimate facts of your life.").

139. See *infra* notes 140, 142 and accompanying text, which discuss the EU's fundamental-rights regime. For a U.S. discussion, see, for example, Elizabeth M. Renieris, Ravi Naik & Jonnie Penn, *You Really Don't Want to Sell Your Data*, SLATE (Apr. 7, 2020, 10:00 AM), https://slate.com/technology/2020/04/sell-your-own-data-bad-idea.html [https://perma.cc/UPZ5-

processing obligations) affords universal and inalienable rights over personal information and enshrines data protection and privacy as fundamental rights.[140] Advocates of this approach in the EU and beyond argue for extending the human-rights framework to data governance as a way to strengthen fundamental data protection in law.[141] Fundamental rights provide individuals with inalienable rights of control over their information, including more stringent consent requirements and ongoing rights of access to data for the data subject.[142] Granting human-rights standing to data subjects would ensure a "minimum standard that cannot be waived by consent, even if all potential uses of data could be foreseen."[143]

Y57F], which argues against a propertarian view of data and advocates for a fundamental-rights framework instead; and Bedoya, *supra* note 138.

140. *See* Charter of Fundamental Rights of the European Union, arts. 7-8, 2012 O.J. (C 326) 391, 397; GDPR, *supra* note 132. The EU's GDPR includes more than just dignitarian data reform. Alongside its suite of individual rights, the GDPR includes a number of affirmative data-processing obligations that apply to data processors regardless of individual consumer choices, and affirmatively requires a lawful basis for any data processing to occur (individual consent is one of six). While there is considerable debate regarding the scope of the GDPR's lawful bases, and how they interact with individual consent and the individual right to restrict processing and the right to erasure, at a minimum, they provide a legal framework for data protection beyond individual ordering. For a helpful explainer, see *Guide to the UK General Data Protection Regulation (UK GDPR)*, INFO. COMM'R'S OFF. (Jan. 21, 2021), https: //ico.org.uk /for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr [https://perma.cc/LJY2-9DA4]. On the debate over the legitimate bases for processing, see Jeff Ausloos, Michael Veale & René Mahieu, *Getting Data Subject Rights Right*, 10 J. INTELL. PROP. INFO. TECH. & E-COM. L. 283 (2019); and Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189 (2019).

141. *See, e.g.*, Elizabeth M. Renieris & Dazza Greenwood, *Do We Really Want to "Sell" Ourselves? The Risks of a Property Law Paradigm for Personal Data Ownership*, MEDIUM (Sept. 23, 2018), https: //medium.com/@hackylawyER/do-we-really-want-to-sell-ourselves-the-risks-of-a-property-law-paradigm-for-data-ownership-b217e42edffa [https://perma.cc/96J6-DTG6].

142. The GDPR grants individuals the following rights over data: the right to be informed, the right of access, the right to rectification, the right to erasure, the right to restrict processing, the right to data portability, the right to object, and rights in relation to automated decision-making and profiling. All of these rights are subject to some overriding exceptions and are undergoing active interpretation in EU law. GDPR, *supra* note 132, arts. 13-22.

143. *The EU General Data Protection Regulation*, HUM. RTS. WATCH (June 6, 2018, 5:00 AM EDT), https://www.hrw.org/news/2018/06/06/eu-general-data-protection-regulation [https://perma.cc/45ZR-P4X8]. Other forms of data governance adopt a less fundamental approach to enacting a minimum standard, seeking instead to heighten the duties owed to data subjects by data collectors in virtue of data's capacity to enduringly and significantly affect the data subject. For instance, several scholars argue for theories of fiduciary obligation, or extend helpful theories of separability from property theory, to individual data governance. *See, e.g.*, Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183

Dignitarian reforms posit a robust legal solution to the problems of data extraction: they enhance the protection of data *about* the subject by making these protections more akin to those afforded the subject *herself*. Dignitarian reforms therefore aim to encode data with features more like those afforded a natural person. On the basis of this quasi personhood, these reforms would extend inalienable rights and impose on others certain duties that ensure personal data is granted a legal baseline of civil and political status. This would formally abolish the quasi-ownership claims to data and instead recognize data's quasi-personhood status, subject to the range of civil-libertarian protections afforded individuals in public life.

Many dignitarian reformers claim that data extraction involves not only individual stakes, but also societal ones. For example, Zuboff says the world's digital information is a public good,[144] and the EU Data Protection Supervisor notes that privacy is "not only an individual right but also a social value."[145] Yet in practice, the legal solutions advanced under dignitarian conceptions of data governance still subject data to individual ordering and protect data subjects from individualist, informational harm.[146] Dignitarian reforms secure negative rights for data subjects against certain downstream uses (for example, use without consent, use that goes beyond the purposes originally given, or use once consent has been withdrawn), and that obtain with respect to data collected about them.[147]

(2016); Richards & Hartzog, *Taking Trust Seriously*, *supra* note 40; Hartzog & Richards, *Privacy's Trust Gap*, *supra* note 40; Verstraete, *supra* note 131. Such theories of reform, while sharing certain relevant features with dignitarian approaches, are not as directly grounded in the dignitarian normative basis for reform and thus warrant separate analysis beyond the scope of this Feature.

144. Alvin Powell, *An Awakening over Data Privacy*, HARV. GAZETTE (Feb. 27, 2020), https://news.harvard.edu/gazette/story/2020/02/surveillance-capitalism-author-sees-data-privacy-awakening [https: //perma.cc/DN69-ABDX].

145. *Data Protection*, EUR. DATA PROT. SUPERVISOR, https://edps.europa.eu/data-protection/data-protectionen [https://perma.cc/W7N5-W85M].

146. Mark Scott, Laurens Cerulus & Steven Overly, *How Silicon Valley Gamed Europe's Privacy Rules*, POLITICO (May 22, 2019, 10:40 AM), https://www.politico.eu/article/europe-data-protection-gdpr-general-data-protection-regulation-facebook-google [https://perma.cc/75YU-YGVW] (noting that despite being previously banned, Facebook's facial-recognition technology is once again permitted in Europe because users are "given the choice to opt into the service" under the consent rules of the GDPR).

147. Under Europe's GDPR, these dignitarian rights are accompanied by a series of affirmative obligations imposed on data processors regarding the storage, transmission, and processing of data. *See* GDPR, *supra* note 132, arts. 24-43 (explaining the duties of a controller and processor); *id.* arts. 44-50 (explaining transfers to third countries)*; see also* Meg Leta Jones & Margot E. Kaminski, *An American's Guide to the GDPR*, 98 DENV. L. REV. 93, 116 (2020) ("In sum, the GDPR consists of two approaches to data protection: a set of individual rights and a set of company obligations."). Whether these affirmative obligations are sufficient to accord

C. Conceptual Limitations of DIM Reforms

While dignitarian reforms offer a more robust individualist regime for data protection than propertarian reforms, like propertarian reforms, they still conceive of data as an individual medium (DIM). As a result, both propertarian and dignitarian reforms attempt to reduce legal interests in information to individualist claims subject to individualist remedies that are structurally incapable of representing the population-level interests that arise due to data-horizontal relations. This fails to fully account for significant forms of social informational harm, and risks foreclosing socially beneficial forms of data production. In this Section, I discuss various conceptual limitations of DIM reforms.

1. Absence of Horizontal Relations

Failing to account for these horizonal relations presents a problem for DIM reforms on their own terms. Using shared population features derived from data about Adam to act upon Ben is what makes such data collection so desirable. This relationality is part of why data collectors face such strong incentives to extract continual data streams from data subjects like Adam to begin with. Horizontal relations, whether explicitly accounted for or not, motivate data collectors to engage in such continual and fine-grained data extraction. Ignoring the interests that result from horizontal relations therefore not only sidelines Ben's interests in such data, but also fails to account for structural conditions that influence the terms of exchange between Adam and TattooView AI, and that in turn index many of the interests that Adam also has in the information collected from him.

Consider again the prior scenario involving Adam, TattooView AI, and Ben. Propertarian reforms would require that TattooView AI pay Adam for his data. Yet payment at the point of collection for Adam does nothing to address how his data is used to detain Ben. Ben incurs significant harm, but receives none of the benefit from propertarian data reforms. In granting Adam the right to payment, propertarian reforms seek to rebalance the terms of Adam's vertical relation with TattooView. Such reforms may ameliorate the worst excesses of data-subject exploitation (and result in some degree of redistribution), but in failing to apprehend both Ben and Adam's legal interests that accrue along horizontal relations, they do not grant Adam (or Ben) the ability to address the conditions structuring the terms of this exchange. Given the practical realities discussed above, such

European data subjects more than individualist protections is a subject of active scholarly debate. *See, e.g.*, Ausloos et al., *supra* note 140; Kaminski, *supra* note 140.

reforms are highly unlikely to produce more equal data relations along either axis.[148]

Dignitarian reforms would admirably extend protection to downstream uses that violate Adam's protected interests in data collected from him. Extending fundamental protections to Adam grants him standing to argue that use of his data to detain him violates his fundamental rights, or alternatively may grant him stronger up-front rights to refuse collection.[149] Yet similar to propertarian reforms, dignitarian rights leave third parties like Ben unaccounted for. Granting Adam rights against having his own data used against him does not affirmatively prevent Adam's data — or the category of tattoo-image data generally — from being used against *others* like Ben for purposes of detention. And yet presumably the interest Ben and Adam have in this information is the same (that is, an interest against having tattoo-image data about their tattoo being used to classify them as suspected gang members and detain them on the basis of this classification). Granting legal protection to one while excluding the other is arbitrary and nonsensical. In both instances, the relevant set of legal interests in this data flow does not reduce to the individual rights granted to Adam by DIM reforms.

Like propertarian reforms, dignitarian reforms fail to apprehend the structural conditions driving the behavior they aim to address. In granting Adam inalienable rights over the terms of his data collection and use, dignitarian reforms seek to rebalance the terms of Adam's vertical relation with TattooView. Dignitarian reforms may ameliorate some forms of data-subject violation. But in failing to index the many horizontal interests at stake, they fail to account for the role horizontal relations play in the economic imperatives of data extraction, as well as the forms of social informational harm such relations may materialize. The observation that data production may violate individual autonomy does nothing to further our understanding of why or how this violation has become an imperative of competitive-market behavior in the data-political economy.[150] Acting on this observation with attempts to strengthen rights of individual data-subject control is thus unlikely to address the structural conditions driving this state of affairs.

---

148. For a more detailed treatment of these conditions, see Salomé Viljoen, *Data as Property?*, PHENOMENAL WORLD (Oct. 16, 2020), https://phenomenalworld.org/analysis/data-as-property [https://perma.cc/QA9X-2585].

149. The merits of such a case are unclear, and beyond the scope of this analysis. For instance, Adam's claim would still be susceptible to all the usual (and in the case of information claims, evolving) constraints of standing analysis. *See* TransUnion L.L.C. v. Ramirez, 141 S. Ct. 2190 (2021).

150. For example, the GDPR does not outlaw the advertising-driven business model that predominantly drives datafication; it requires companies to be more transparent about this use and gives users greater access to how their data is being used.

2. *Missing or Misdiagnosed Theories of Harm*

The absence of horizontal data relations in law may cause data-governance law to miss — or misconceive of — how data production results in particular kinds of injustice. As detailed above, datafication gives rise to two classes of critique or claims of injustice: the inequality diagnosis and the commodification diagnosis. The *inequality diagnosis* locates the injustice of data production in the unfair distribution of wealth that datafication creates. It conceives of the injustice of datafication as one of unjust enrichment. The *commodification diagnosis* locates the injustice of datafication in the excessive legibility of data subjects that results. This diagnosis conceives of the injustice of datafication as the wrongful control this excessive legibility grants data collectors over data subjects. This control in turn undermines data-subject autonomy and violates their dignity by reducing their inner lives to transactions mined for value.

These two articulations or diagnoses of what makes datafication wrongful in turn motivate the two DIM agendas for reform. Propertarian reforms aim to respond to the inequality diagnosis by granting data subjects a right to reclaim some portion of the material benefits created from data production. Dignitarian reforms aim to respond to the commodification diagnosis by reasserting greater control for data subjects over if, when, and how they may be rendered legible by data collectors.

Yet each reform fails to respond to the diagnosed injustice of the other. Propertarian reforms by design concede extensive data-subject legibility as a necessary condition for securing some redistributive benefit. Dignitarian reforms by their own commitments cannot provide data subjects with material redistributive value, as this would violate dignitarian prescriptions against commodifying knowledge of the inner self. Even if one assumes each reform can address its diagnosed form of injustice (and as the previous subsection notes, there are significant reasons not to make such an assumption), choosing one leaves the other diagnosis of injustice unaddressed. If one believes both reforms capture compelling concerns regarding data production, then pursuing the either/or path of DIM reforms presents a dilemma.

3. *Unjust Data Production as Unequal Data Relations*

Each diagnosis and related agenda for reform presents both a normative issue (i.e., not addressing a valid aspect of what makes datafication wrongful), as well as an operational issue (i.e., missing relevant features in its attempt to address its own diagnosis of what makes datafication wrongful) that leaves each type of reform unlikely to materially address the problems motivating reform.

Reconceptualizing these diagnoses of injustice to account for data relations may resolve these issues in helpful and clarifying ways. What makes datafication wrongful is neither that it represents unjust enrichment nor that it is an instance of wrongful self-commodification. Datafication (or, more precisely, data production) is wrongful if and when it materializes unjust social relations along either the vertical or horizontal axis. These unjust social relations may take the form of exploitative data relations that generate unfair wealth distributions, as well as data relations that materialize forms of group oppression like racism, sexism, and xenophobia. By centering data relations in our diagnosis of injustice, we can recast the reform agenda of data-governance law as managing (and ideally equalizing) these data relations.

This alternative normative diagnosis also helps pinpoint what DIM legal agendas miss. Data production's role in enacting or amplifying inequality is not simply a matter of data-subject nonpayment, but concerns the unjust social relations being amplified or enacted on the basis of shared population features. Payment at the point of collection may redistribute some portion of the profit that results from exploitative data collection, but does nothing to address how data production itself may amplify or enact social oppression as a means to generate that profit.[151] Even if payment were to distribute the gains from data production in a completely egalitarian manner, datafication as a process materializing unequal and oppressive social relations would remain.

The focus on social inequality (as opposed to unjust enrichment) also captures relevant aspects of dignitarian concerns regarding algorithmic governmentality. Governmentality via data-driven feedback systems is wrong not only because it undermines processes of self-formation (though it may well have this effect), but also because such systems enact unjust social relations that serve to dominate, marginalize, and demean.[152] Recasting the injustice of surveillance

---

151. In fact, by legitimating the marketplace for data, payment may serve to legitimate downstream practices that result from lawful engagement in that marketplace. Because data is commoditized to begin with, ICE was able to purchase access to this database from its provider, Venntel, as opposed to gathering this data itself. This commercial exchange provides ICE strong legal protection for using this data. Under *Carpenter v. United States*, ICE may have needed a warrant to obtain this data from carriers or app companies directly. 138 S. Ct. 2206, 2209 (2018). Yet because ICE simply purchased access to the database from a data broker, as could any other entity, any potential constitutional challenge is weakened.

152. For example, consider the growing literature on how algorithmic forms of self-knowing enact cultural imperialism. *See, e.g.*, COULDRY & MEJIAS, *supra* note 33, at 5-6; Dan M. Kotliar, *Data Orientalism: On the Algorithmic Construction of the Non-Western Other*, 49 THEORY & SOC'Y 919, 922-25 (2020). Cultural imperialism refers to the universalization of a dominant group's experience or culture and its establishment as the norm. This grants the dominant group primary access to what Nancy Fraser calls the "means of interpretation and communication" in

data flows as that of unequal social relations brings into view the structural forces driving personal instances of violation as well as the mutual stakes we have in the injustice of such conditions.

Recasting data-governance reform as equalizing data relations also helpfully clarifies a distinction glossed over in dignitarian accounts between "commodification" and "legibility" regarding what makes legibility wrong: namely, the goals motivating apprehension and the substantive and procedural conditions that determine those goals. This distinction vanishes in critiques against private companies like Facebook, which are currently the subject of the fiercest dignitarian critiques. But it is relevant for distinguishing the data collection and use of private companies from those of publicly or otherwise collectively accountable data infrastructures.

The relevant inquiry is not whether and to what degree a data subject has been rendered legible to a given system (and whether they had the opportunity to exert control over this process), but to what ends and under what conditions such legibility occurs—and most importantly, whether these have been determined in ways that enact more equal data relations. Under this account, permissible legibility is not simply a matter of individual data-subject consent or control, but one of the institutional forms that adjudicate between and determine the legitimate and illegitimate bases for data production.

Consider again the example of TattooView AI collecting user data to detain suspected gang members. What makes the tattoo-data flow potentially unjust is not that the population at the point of data collection was not paid, but that information about one group (the data subjects) is being used to oppress and dominate others on the basis of their ascribed group membership (that is, "gang member," a group membership informed by racial, ethnic, class, and linguistic differences). This tattoo-data flow is not (only) unjust because its collection or its use renders Adam legible in ways that may violate Adam's autonomy and his right to self-determination. It also materializes a social category ("gang member") that, when acted upon, results in the domination and oppression of others. Under propertarian and dignitarian reforms this social effect continues to have no bearing on how information law regulates what data may be collected, stored, exchanged, or used.

---

a society. NANCY FRASER, SOCIAL MOVEMENTS VS. DISCIPLINARY BUREAUCRACIES: THE DISCOURSES OF SOCIAL NEEDS 7-8 (1987). Often without realizing it, dominant groups project their experiences as the experiences of humanity; the result is cultural products of communication and sense-making that reflect dominant experiences, values, goals, and achievements. This creates the culturally oppressed experience that W.E.B. Du Bois called "double consciousness": the sense of "always looking at one's self through the eyes of others, of measuring one's soul by the tape of a world that looks on in amused contempt and pity." W.E. BURGHARDT DU BOIS, THE SOULS OF BLACK FOLK 45 (NAL Penguin, Inc. 1969) (1903).

4. DIM Reforms and Socially Beneficial Data Production

Reducing interests in the digital economy to individual data-subject interests may inadvertently foreclose socially beneficial forms of data production. Currently, a predominant purpose that draws critiques of datafication is that of private wealth creation.[153] Wealth creation is one purpose of collecting data, but there are others. For example, running social-welfare enterprises that require at-scale distribution and management of precious resources (such as water) or that require time-sensitive predictions for overriding public interests (such as public-health strategies to limit the spread of COVID-19) themselves require high-quality population data to ensure public-welfare obligations are met effectively and fairly.

Yet the diagnoses of harm under DIM reforms do not index these distinctions, and neither do the legal agendas that result from them. Under these accounts, datafication for the public interest and datafication for private-wealth creation pose the same risk of individual violation and are subject to the same forms of individualized governance. Under propertarian regimes, if a public agency cannot pay data subjects a fair price for this data, it may well be argued that such datafication constitutes a public taking or should be subject to individual decisions to donate such data or not. Under dignitarian regimes, an individual may disagree with the public purpose (e.g., they might believe that government efforts to trace COVID-19 violate their medical liberty) and deny access to their data on the basis that this use violates their individual will and their fundamental rights. Both instances — taking data for free or collecting it absent consent and for a purpose the data subject disagrees with — violate individualist conceptions of how information's collection and use should be ordered, and what conditions of datafication are legitimate. Yet relying on individuals to participate

153. Another use of datafication that draws critique is law enforcement and government surveillance. In the context of the United States, however, most high-profile scandals regarding law enforcement's use of technology involve private entities selling surveillance products to law enforcement. This structure, again, is one particular business model under the organizing principle of datafication for the purpose of private wealth creation, which fuels the ubiquity of personal data-based surveillance products available for sale on the private market. Lyons, *supra* note 103; Dana Goodyear, *Can the Manufacturer of Tasers Provide the Answer to Police Abuse?*, NEW YORKER (Aug. 20, 2018), https://www.newyorker.com/magazine/2018/08/27/can-the-manufacturer-of-tasers-provide-the-answer-to-police-abuse [https://perma.cc/W5T3-5WZX].

voluntarily in these systems may significantly undercut their capacity to realize the broader social benefits they are meant to achieve.[154]

DIM reforms thus suffer from being simultaneously overly narrow and overly broad. By focusing on datafication's violation of self- or uncompensated-value creation, they do not address the economic imperatives that drive such harm nor do they provide an effective agenda for addressing inequality in the data-political economy. At the same time, the focus on datafication paints over meaningful distinctions between the purposes of data production and the conditions under which such purposes are determined.

IV. DATA AS A DEMOCRATIC MEDIUM

A. The Legitimacy Problem

Both current and proposed individual-level rights in data cannot address the population-level interests that arise from data production. As a result, these reforms are unable to resolve the legitimacy problem that, alongside the sociality problem, continues to vex U.S. data-governance law. The legitimacy problem asks: how can data-governance law distinguish legitimate from illegitimate data use without relying on individual adjudication?

Consider the following example. Suppose that an entity (Watercorp) is collecting data on household water consumption. Every time someone drinks water from the tap, sets their kettle to boil, waters their herb garden, or brushes their teeth, that information is collected, processed, and analyzed, with the goal of changing future water-usage behavior for the households in a given municipality. This data reveals intimate facts about people's lives. Indeed, from this data emerges a detailed portrait of their daily habits. This detailed portrait may be used for any number of reasons: to help households set and meet water reduction goals; to calculate "surge" prices for water usage based on peak consumption; to use feedback data to shift people's water-consumption patterns toward

154. For instance, public-health authorities deploying COVID-19 digital contact-tracing apps targeted a sixty-percent population threshold for the systems to work most effectively in counteracting the pandemic. Although lower numbers of app users are still estimated to reduce the number of coronavirus cases, getting closer to the sixty-percent threshold significantly increases the efficacy of digital-tracing systems. *Digital Contract Tracing Can Slow or Even Stop Coronavirus Transmission and Ease Us Out of Lockdown*, U. OXFORD RSCH. (Apr. 16, 2020), https://www.research.ox.ac.uk/Article/2020-04-16-digital-contact-tracing-can-slow-or-even-stop-coronavirus-transmission-and-ease-us-out-of-lockdown [https://perma.cc/8NUL-DJ88]. *But see* Patrick Howell O'Neill, *No, Coronavirus Apps Don't Need 60% Adoption to Be Effective*, MIT TECH. REV. (June 5, 2020), https://www.technologyreview.com/2020/06/05/1002775/covid-apps-effective-at-less-than-60-percent-download [https://perma.cc/Y33S-CDNP] (arguing that even at a much lower percentage of users, apps are still effective).

the bottled drinks of a client; to sell insights about people's daily habits to advertisers, insurers, creditors, and employers; and to test what strategies make people most likely to pay their utility bills.[155]

Now suppose instead of Watercorp, Waterorg—the municipal public authority for the drought-prone area—engages in this same sort of data collection to understand the municipality's water usage, develop strategies to reduce water consumption, and (as droughts grow more severe) develop plans to ensure water will be distributed fairly and responsibly as it becomes scarcer. Suppose further that the risks of drought-based water shortages and shutoffs are highest in the driest and poorest districts of the municipality, where a higher proportion of minority residents live.[156]

Finally, suppose that a handful of citizens of the municipality object to the coercive power of the state in collecting this data from them or to using this data to inform water-allocation strategies that affect them. They argue that this data collection violates their dignity and autonomy. It extracts their intimate water-consumption data against their own interests, since it will result in future water policies that will almost certainly reduce their capacity to freely access and use water as they choose, free from observation. Like Watercorp, Waterorg is exerting coercive power to render such citizens legible to Waterorg against their will, without payment, and for purposes that go against their interests.

If one is concerned about Watercorp's data production, but believes that it is permissible or even responsible for Waterorg to engage in the same kind of data

155. A recent, controversial, and randomized controlled trial ran an experiment to see whether shutting off tenants' water makes landlords more likely to pay their water bills. *See* Josh Budlender (@JoshBudlender), TWITTER (Aug. 8, 2020, 2:48 PM), https://twitter.com/joshbudlender/status/1292170843389386761 [https://perma.cc/FDP6-LWRV] (displaying screenshots of the paper, which has been temporarily withdrawn).

156. Though at a smaller scale, this distribution is not unrealistic. Evidence suggests that water stress as a result of a changing climate will disproportionately impact poorer communities. The World Health Organization estimates that "[b]y 2025, half of the world's population will be living in water-stressed areas" and one-quarter of the world's population currently faces "extremely high" levels of water stress. *Drinking-Water*, WORLD HEALTH ORG. (June 14, 2019), https://www.who.int/news-room/fact-sheets/detail/drinking-water [https://perma.cc/8XZC-WVB8]; Rutger Willem Hofste, Paul Reig & Leah Schleifer, *17 Countries, Home to One-Quarter of the World's Population, Face Extremely High Water Stress*, WORLD RES. INST. (Aug. 6, 2019), https://www.wri.org/blog/2019/08/17-countries-home-one-quarter-world-population-face-extremely-high-water-stress [https://perma.cc/7GBK-U8S9]. North Africa and the Middle East contain twelve of the seventeen most water-stressed countries; India ranks thirteenth internationally for water stress and has more than three times the population of the other sixteen most stressed countries combined. *Id.* In the United States, New Mexico faces extreme water stress, and California, Arizona, Colorado, and Nebraska all face high water stress. *Id.*

production for different purposes, analysis under DIM accounts of data governance presents a challenge. Waterorg's basic governance structure allows for broader, democratic representation in the determination of societal goals—accompanied by constitutional constraints against certain forms of individualized use. For Watercorp, we have no means for making democratic decisions about the collective societal goals, nor are there the same constitutional forms of oversight that serve as substantive backstops against impermissible collection and use. Focusing on the preferences or rights of each individual or household regarding whether to participate in this collection does not apprehend the normatively distinct purposes and conditions of data production from these two entities. Further, this approach fails to accord relevance to the mutual and overlapping interests these households have in one another's choices.

B. Horizontal Relations and Institutional Design

To address the relevant distinctions between Watercorp and Waterorg's data production and to adjudicate their legitimacy requires recourse to population-level, democratic evaluation of these proposed data-production schemes.

Under DIM, an evaluation of the legitimacy of such a scheme would lead to the various inquiries. Did these households adequately consent to this tracking? Do the purposes for which this water data is being used uphold the rights of household members? Do they violate any duties owed to the household members? Alternatively, are households being adequately compensated for this data collection? In response to the citizens who object to data being collected, robust DIM reforms would grant citizens the ability to deny collection, the right not to have data about them used in ways that violate their interests, or payment for the data they choose provide.

Under this analysis, both Waterorg and Watercorp's behavior may be diagnosed as wrongful (and, if addressed via legal reform, unlawful) if either entity collects and commodifies household-water data against the wishes or interests of households from whom it is collected. If household members feel wrongfully commodified, under robust dignitarian DIM protections, they would have the right to object to and opt out of this data production; under robust propertarian DIM protections, they would have a right to demand a greater share of the wealth their data creates for Watercorp or fair repayment under takings law from Waterorg. On the other hand, Waterorg or Watercorp's behavior under this analysis is not wrongful or unlawful if they collect this data under robust conditions of meaningful consent, do not use this data in ways that violate the protected legal interests of household individuals, provide real options for households to opt out of water collection, or, alternatively, provide a fair wage or sale price for

household-water data. In sum, these protections, done right, secure for households the rights to payment, exit, or recourse, regardless of which entity is collecting their data or which purposes guide this collection. This approach may empower individual households against either entity, but it still practically falls back on individual choice to determine the legitimacy of data collection.

Even robust DIM-based responses miss how population-level interests in data production work. Consider the citizens who object to Waterorg's data collection due to the possible adverse use of such data against them (let's call one such citizen "Cate"). Cate's concern over the adverse use of household-water data neither reduces to a right to prevent such data from being collected from her home nor to a right to restrict how data from her home may be used. Instead, her concern presents a population-level interest in all household-water data. Waterorg doesn't need her data to get population-level insights about water consumption habits for households like hers—they may easily derive such insights from households that share relevant features (for example, same household size, same neighborhood). For Cate's concern to be effectively expressed, it would need to be accounted for at the population level: for municipal water-data production as a whole.

Nor is Cate's interest the sole interest at stake in water-usage data collected from her home. Without enough quality household-water data, Waterorg may not be able to make sufficiently fair or accurate water-allocation plans as droughts grow more severe. This stymies Waterorg's plans to develop drought-conscious water management not just for those who withheld their data, but for everyone in the municipality. Fair and effective water management is particularly important for those who live in the poorer, drought-prone areas. The risks of noncollection will fall disproportionately on them, amplifying the material hardship experienced by the community that lives there.[157] These interests also accrue at the population level for water-data production as a whole.

Because data subjects' interests do not reduce down to how their individual water-consumption data is extracted or used, and because there are multiple interests at stake beyond that of the individual data subject, it is impossible to determine the legitimacy of Waterorg's data production simply by referring to the conditions of data subjects' interpersonal exchanges with Waterorg. Only by recognizing the full array of interests that are relevant to the task of governance can we begin to address the forms of social informational harm that may arise as a result of them.

---

157. Note that this example explores a positive purpose for data collection (water allocation) that stands to disproportionately benefit this poorer community. One can also imagine a negative example that may produce disproportionate risks to this poorer community and would give rise to an interest for this community in noncollection. But again, this interest in data production would obtain at the institutional level concerning *all* water-collection data.

C. *Democratic Data Governance*

Reconceptualizing the project of data governance from that of securing individual rights to institutionalizing collective ordering, shifts the relevant line of inquiry. In the first instance, the question is how to secure greater data-subject control or better legal expressions of data-subject autonomy. The new line of inquiry asks how we can balance the overlapping and, at times, competing interests that comprise the population-level effects of data production. This reorientation raises core questions of democratic governance: how to grant people a say in the social processes of their own formation; how to balance fair recognition with special concern for certain minority interests; how to identify the relevant "public" or institutional level of civic life at which to coalesce and govern such collective interests; and how to not only recognize that data production produces winners and losers, but also develop fair institutional responses to these effects.

This shift, in turn, theorizes a different approach to data in law—from an individual medium expressing individual interests, to a democratic medium that materializes population-level, social interests. Like other mediums of social relation, the governance of data raises political questions regarding what individuals are owed and owe one another on the basis of these material relations, and how to distribute relevant benefits and risks among one another. This conceptualization of data is referred to below as "data as democratic medium" (DDM).[158]

1. *Democracy as a Normative (Egalitarian) Standard*

Asserting that data relations are "democratic" is to take an additional step beyond the descriptive claim that data is "relational." This assertion expresses distinctly political and normative criteria for how data's relationality and its attendant social effects should be negotiated and managed.[159] Conceptualizing data as a democratic medium therefore incorporates both a positive and a normative claim: describing the kinds of interests that do result from data production as well as how such interests should be governed. Put differently, the fact

158. The author wishes to credit a conference held by Christine Desan, "Money as a Democratic Medium," at Harvard Law School in December 2018 for inspiring this phrase's application in the data-economy context. *See* Brette Milano, *Money as a Democratic Medium*, HARV. L. TODAY (Jan. 11, 2019), https://today.law.harvard.edu/money-as-a-democratic-medium [https://perma.cc/7A7P-DMVP].

159. *See, e.g.*, KASPER LIPPERT-RASMUSSEN, RELATIONAL EGALITARIANISM: LIVING AS EQUALS 27 (2018); Samuel Scheffler, *The Practice of Equality*, *in* SOCIAL EQUALITY: ON WHAT IT MEANS TO BE EQUALS 21, 31 (Carina Fourie, Fabian Schuppert & Ivo Wallimann-Helmer eds., 2015) (arguing that equality is a form of practice rather than a normative pattern of distribution); IRIS MARION YOUNG, JUSTICE AND THE POLITICS OF DIFFERENCE 16 (2011) ("[D]istributive paradigm is a tendency to conceive social justice and distribution as coextensive concepts.").

that data is relational (and gives rise to irreducibly social interests) implies that data should thus be governed democratically—a mode of governance that gives moral and legal force to data's relationality.

Democracy as a normative standard offers criteria for evaluating how data relations are ordered, and should be ordered, by data-governance law. It provides one theory of what features define unjust data relations and distinguish them from just relations. "Thorough social and political democracy," writes Iris Marion Young, "is the opposite of domination."[160] Democratic equality is achieved, argues Elizabeth Anderson, under conditions in which "people stand in relation of equality to others."[161] Developing democratic institutions whereby people relate as equals does not merely secure the social conditions of individual freedom; it also addresses the institutional arrangements by which people's opportunities are generated over time and "reflects a deontic requirement grounded in our equal moral status as persons."[162] Institutional recognition of competing interests therefore operationalizes the normative force of the population-level effects that one's personal choices over data have on others and may express not only what individuals are owed, but also what their obligations are to one another.[163] This framing posits an egalitarian-political standard for legitimacy in place of individual choice, which considers the quality of relations under which data production occurs and those it seeks to enact.

Democratic ordering can therefore also provide one substantive standard by which to evaluate and distinguish different goals of data production, on the basis of the goals it seeks to achieve and the social relations under which production occurs. In the context of data production, the general egalitarian case for democratic ordering is bolstered by the specific, empirical significance of population-level interests in data production. DDM therefore expresses not only the general political case in favor of more democratic ordering, but also something akin to empirical fact: personal choices over data sharing should reflect the effects these choices have on others, not only because of the political and moral benefits of considering others, but also because under current conditions of datafication, individuals already directly relay information relating to others, which is used to predict and influence the behavior of others.

---

160. YOUNG, *supra* note 159, at 38.

161. Elizabeth Anderson, *What Is the Point of Equality*, 109 ETHICS 287, 289 (1999); *see also* Elizabeth Anderson, *Toward a Non-Ideal, Relational Methodology for Political Philosophy: Comments on Schwartzman's* Challenging Liberalism, 24 HYPATIA 130 (2009) [hereinafter Anderson, *Comments*].

162. LIPPERT-RASMUSSEN, *supra* note 159, at 19; *see* Anderson, *Comments*, *supra* note 161.

163. LIPPERT-RASMUSSEN, *supra* note 159, at 24-25; Scheffler, *supra* note 159, at 38-39.

2. *Democratic Evaluation of Waterorg vs. Watercorp*

On this view, we can provide a more precise and complete account of whether—and why—Watercorp's data production is wrongful. The problem with Watercorp's data production is not that Watercorp extracts household data without consent, or underpays households, or renders households legible against their will. Instead, Watercorp's data production suffers from a more fundamental problem: that households under Watercorp's data production scheme have no ability to meaningfully determine the quality of the data relations they are being placed in—the social processes via data production to which they are being subjected. They cannot exercise equal power either back onto Watercorp or over one another with respect to the population-level decisions that affect them all. In other words, under the current legal arrangement, Watercorp does not have to consider the normative force of its decisions or actions on others—and neither do individual households who may choose to opt in or out of this data production.[164] Under the Watercorp scheme, households have (at best) an incomplete say in the institutional arrangements that structure the scope of their choices and the social processes to which they are subjected. Securing negative rights of exit or payment are not the same as securing affirmative rights to representation in the conditions and purposes of data production.

This account clarifies that Waterorg's data production may be legitimate even if it subjects data subjects to mandatory data collection, so long as the fundamental condition of full institutional recognition is satisfied. What population-level interests make clear is that the relevant task of data governance is not to reassert individual control over the terms of one's own datafication (even if this were possible) nor to maximize personal gain, but instead to develop the institutional responses necessary to represent the relevant population-level interests at stake in data production. This shifts the task of reform from providing opportunities for exit, payment, or recourse, to securing recognition and standing to shape the purposes and conditions of data production, thus establishing the terms of legitimate mutual obligation.[165]

---

164. *See* Scheffler, *supra* note 159, at 25. Scheffler identifies the relational egalitarian ideal as a deliberative constraint: people are in a relation of equality where each person accepts that the other person's equally important interests should play a mutually significant role in influencing the decisions that govern that relationship—that each person's "equally important interests constrain [their] joint decisions to the same extent." *Id.*

165. *See* NANCY FRASER, JUSTICE INTERRUPTUS 11-39 (1997) (exploring recognition as a remedy to injustice). *See generally* AXEL HONNETH, THE STRUGGLE FOR RECOGNITION 1 (Joel Anderson trans., Polity Press 1995) (1992) (setting forth, "on the basis of Hegel's model of a 'struggle for recognition', the foundations for a social theory with normative content").

Population-level representation also clarifies the trade-offs among competing interests in data production. In the Waterorg scheme, shifting from individual rights to institutional governance represents both the interests of the citizens who oppose data collection and the interests of citizens who stand to suffer the most without such collection. This clarifies who stands to lose and who stands to gain from data production, as well as the potentially distinct normative stakes of their relative wins and losses.

D. Benefits of DDM

As the previous hypothetical shows, reconceptualizing the relevant interests at stake in data governance to account for data relations has several benefits. It indexes the social injustice of datafication and places social informational harm on equal footing with individual-informational harm. It also provides a theoretical basis from which to make the positive case for vital forms of social-data production. Like analog social relations, data relations may be (and all too often are) oppressive or exploitative; but like analog social relations, data relations may be empowering, enabling people to achieve social goals together that they cannot accomplish alone. These benefits, as well as a few insights regarding how DDM relates to the interests that animate individualistic accounts, are discussed below.

1. Social Informational Harm

Reconceptualizing what interests are relevant for data governance clarifies what makes data production, as a core economic activity in the digital economy, potentially wrongful. Data production may indeed be unjust if data subjects are manipulated at the point of collection or subject to governmentality at the point of use. Such acts may wrongfully violate data-subject autonomy. But data production may *also* be unjust when it enacts or amplifies social processes of oppression along horizonal data relations. Indeed, evidence suggests that this is a large and growing problem in the digital economy, and a significant source of the social and political critiques levied against large data producers.[166]

As an unjust social process, datafication denies individuals (both data subjects and those with whom they are in horizontal-data relations) a say in the

166. This point is covered in some detail in the Introduction, *supra*. *See also* Salomé Viljoen, The Promise and Limits of Lawfulness: Inequality, Law, and the Techlash (Nov. 23, 2020) (unpublished manuscript), https://ssrn.com/abstract=3725645 [https://perma.cc/8AWJ-BB9T].

social processes of their mutual formation. In so doing, data relations can materialize unjust group-based relations like racism, sexism, and classism.[167]

Let us take again the example of ICE detaining undocumented immigrants on the basis of their movement patterns. Movement patterns as a shared feature become one defining feature of the category of "undocumented immigrant" (a category which in turn is defined via racial, class, and linguistic difference). By identifying this common feature and operationalizing it to detain people, this data flow materializes a particular oppressive social meaning onto the category of "undocumented immigrant." Such data flows thus become social fibers of domination: they help to create, organize, express, and direct the meaning of this social category as the experience of systematic violence and oppression for those who occupy this category.[168] This gives social meaning to the category of "undocumented immigrant," such that part of what group membership becomes is the fact of having the movement patterns of yourself and others weaponized against you.[169]

This form of injustice is a fellow traveler of personal violation—it denies individual undocumented immigrants the chance to determine their own social formation—but it also represents a distinct form of social injustice. It structures

167. This theory of injustice is far from new. Several political philosophers and legal theorists, including those whose analysis informs my own, such as G.W.F. Hegel, Nancy Fraser, Elizabeth Anderson, Axel Honneth, and Samuel Scheffler, view social relations as the (or a) primary basis of (in)justice. This view also builds on social-constructivist accounts of group membership; these accounts center the social meaning of group membership—the cultural practices, institutions, norms, and material conditions that make group membership coherent indicators of identity and experience, and for relevant forms of group membership (race, gender, caste, nationality, etc.) also define forms of oppression that attend (and constitute) group membership. *See, e.g.*, SALLY HASLANGER, RESISTING REALITY: SOCIAL CONSTRUCTION AND SOCIAL CRITIQUE 3-32 (2012). Haslanger draws on feminist and critical race theory to develop the idea that gender and race are positions within a structure of social relations. On this interpretation, the point of saying that gender and race are socially constructed is not to make a causal claim about the origins of our concepts of gender and race, or to take a stand in the nature/nurture debate, but to locate these categories within a realist social ontology. This is politically important, for by theorizing how gender and race fit within different structures of social relations, Haslanger argues that we are better able to identify and combat forms of systematic injustice.

168. *See* Catherine A. MacKinnon, *Feminism, Marxism, Method and the State: An Agenda for Theory*, 7 SIGNS 515, 516 (1982).

169. On systematic violence, see YOUNG, *supra* note 159, at 61-62. Young defines a particular form of systematic violence as a system of social oppression. Members of oppressed groups often live with knowledge that they must fear random, unprovoked attacks on the basis of group membership. The social practice of violence serves to reproduce social oppression through its assertion onto the meaning of group identity and its making a feature of group membership the experience of fearing a particular form of violence. Catherine A. MacKinnon famously advances this argument regarding the social construction of sexuality via hierarchical relations of desire. *See* MacKinnon, *supra* note 168.

a hierarchical group relationship between undocumented immigrants and others. DDM's conceptual account thus helpfully identifies why patterns of datafied personal violations reinscribe existing social arrangements of patterned disparity on the basis of race, sex, class, and national origin. Focusing legal inquiry on data production's population-level effects brings into view both how and why the risks of personal violation are not randomly distributed, but determined via existing social patterns of power distribution that occur along the lines of group membership.[170]

In short, by forming and then acting on population-level similarities in oppressive and dominating ways, datafication may materialize classificatory acts of oppressive-category formation that are themselves unjust. This adds a social dimension to the personal violations of governmentality. Datafication is not only unjust because data extraction or resulting datafied governmentality may violate individual autonomy; datafication may also be unjust because it violates ideals of social equality. Social informational harm thus represents an additional and fundamental form of potential injustice of relevance for data-governance law. Locating material forms of social injustice in datafication also helps to identify data production as an important terrain in other debates regarding why social processes that enact group oppression may be wrong, and how they may be apprehended and addressed via law.[171]

---

170. *See* CHARLES TILLY, DURABLE INEQUALITY 8-9 (1998); YOUNG, *supra* note 159, at 40-48 (defending oppression as a condition of social groups, one that designates the "disadvantage and injustice some people suffer" as resulting from the structural position of that social group, rather than the "conscious and intentional oppression of one group by another"); *see also* MARILYN FRYE, *Oppression*, *in* THE POLITICS OF REALITY: ESSAYS IN FEMINIST THEORY 1, 10-11 (1983) ("[Oppression is] an enclosing structure of forces and barriers which tends to the immobilization and reduction of a group or category of people.").

171. *See, e.g.*, Issa Kohler-Hausmann, *Eddie Murphy and the Dangers of Counterfactual Causal Thinking About Detecting Racial Discrimination*, 113 NW. U. L. REV. 1163 (2019) (arguing that the social theory of discrimination underlying the models of discrimination in law and social science is based on a flawed theory of what the category of race references, how it produces effects in the world, and what is meant when we say it is wrong to make decisions of import *because of* race); Lily Hu, *Direct Effects*, PHENOMENAL WORLD (Sept. 25, 2020), https://phenomenalworld.org/analysis/direct-effects [https://perma.cc/GQZ7-CSUA]; MARTHA MINOW, MAKING ALL THE DIFFERENCE (1990); YOUNG, *supra* note 159. The rich and lively debate in political and social philosophy regarding whether properly attending to group membership requires a group-based methodology for identifying features of justice, or a group-based theory of justice, is ongoing. This worthwhile debate is complex and beyond the scope of this piece, which will simply identify here the importance of group membership and the role of category construction in social processes of injustice for many theorists (a few of which are cited above) in understanding how social injustice works, and thus what justice may require for groups *qua* group membership. *See, e.g.*, LISA H. SCHWARTZMAN, CHALLENGING LIBERALISM: FEMINISM AS POLITICAL CRITIQUE 1-4, 8-11 (2006) ("[R]ights should be seen also as 'goals' that need to be sought after and achieved through structural changes in social power structures."); Anderson, *Comments*, *supra* note 161; HASLANGER, *supra* note 167.

2. *Socially Beneficial Data Production*

DDM offers an opportunity to conceptually distinguish purposes and priorities of data production for socially worthwhile ends. In so doing, it offers a robust positive agenda for data-governance law to expand on existing practices of data production for the public interest, undertaken with strong forms of public accountability, purpose limitations, and confidentiality standards.

a. Expanding on Existing Practices

Public data collection and use have long served a key role in the institutional management of state welfare and in other instances of public-knowledge management for public benefit. Public health care information systems like the UK's national health data sets, or the Veterans' Affairs Administration's open-source electronic health-records system VistA, facilitate high-quality public-health research.[172] Statistics on U.S. demographics and economic activity collected by the Bureau of Labor Statistics and other U.S. statistical agencies offer invaluable insight into the changing patterns of American life. The basic task of governance could not be achieved without the massive collection of tax information by the Internal Revenue Service, nor could financial regulation occur without the disclosure requirements overseen by the Securities and Exchange Commission.

Governance with any commitment to public welfare will always require balancing the necessity of collecting important—and at times highly personal and consequential—information from citizenry, and the risk of oppression and undue coercion that accompanies that collection. Yet, as this Feature argues above, the absence of public oversight does not signify the absence of potentially coercive and harmful effects from data production. Indeed, existing best practices and several protodemocratic proposals for data governance offer promising examples of how to achieve robust legal protections against socially harmful data production, while preserving the societal benefits data production may facilitate.

---

172. On the Department of Veterans' Affairs (VA), see PHIL LONGMAN, BEST CARE ANYWHERE 34-43 (2d ed. 2010). *See also* Arthur Allen, *A 40-Year 'Conspiracy' at the VA*, POLITICO (Mar. 19, 2017, 7:56 AM EDT), https://www.politico.com/agenda/story/2017/03/vista-computer-history-va-conspiracy-000367 [https://perma.cc/6EA4-2ABG] (detailing the creation and maintenance of electronic health records at the VA, and characterizing the VA's system as "topping the lists of the most effective and popular medical records systems"). The author wishes to thank and credit Chris Morten for this excellent example. On the national health data sets, see *Data Sets*, NHS (Apr. 1, 2021, 1:42 PM), https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets [https://perma.cc/B959-VQ2X].

There are several protodemocratic data-governance proposals and projects from which to draw inspiration. Several proposals aim to assert public management and control over existing proprietary data flows, often via mandated public access or by reverting such data to the public domain to be managed via public trust.[173] One possibility is for data-governance legislation to require private-data companies to provide national statistical officers (appropriately safeguarded) with access to private-data sets under specifications set by law or agency determination.

A bolder alternative is to build on examples like the Human Genome Project to develop a public-data management authority for public benefit, rather than for proprietary gain. Former German Social Democrat leader Andrea Nahles has argued for a national data trust, likening digital-technology companies to pharmaceutical companies that enjoy a limited monopoly right to their data.[174] After a set period of years, such data would revert to the public domain to be governed by a public trust or independent agency for use in service of the public good.[175] The UK and Canada have explored public-data trusts as a way to collectively govern citizen data as a national resource from which to develop competitive

173. Jathan Sadowski, Salomé Viljoen & Meredith Whittaker, *Everyone Should Decide How Their Digital Data Are Used—Not Just Tech Companies,* NATURE (July 1, 2021), https://www.nature.com/articles/d41586-021-01812-3 [https://perma.cc/WAJ4-96E6].

174. Andrea Nahles, *Die Tech-Riesen des Silicon Valleys Gefährden den Fairen Wettbewerb*, HANDELSBLATT (Aug. 13, 2018, 6:16 AM), https://www.handelsblatt.com/meinung/gastbeitraege/gastkommentar-die-tech-riesen-des-silicon-valleys-gefaehrden-den-fairen-wettbewerb/22900656.html [https://perma.cc/D5Q3-QKS2].

175. *See* Evgeny Morozov, *There Is a Leftwing Way to Challenge Big Tech for Our Data. Here It Is*, GUARDIAN (Aug. 19, 2018, 12:59 AM EDT), https://www.theguardian.com/commentisfree/2018/aug/19/there-is-a-leftwing-way-to-challenge-big-data-here-it-is [https://perma.cc/9CKH-K86J]; Nahles, *supra* note 174; Hetan Shah, *Use Our Personal Data for the Common Good*, 556 NATURE 7 (2018) (arguing in favor of public data governance for the common good).

technology industries.[176] Barcelona has implemented a civic-data trust to manage its data commons, democratizing data governance while also using its data infrastructures to deepen democratic engagement.[177]

Such proposals can be distinguished from individualist-propertarian approaches in that they do not extend individual rights to data subjects as a way to break open the walled gardens of corporate-held consumer data. Instead, they conceive of citizen data as a public resource (or infrastructure) to be managed via public governance and in furtherance of public goals. Such proposals also depart from dignitarian approaches; they advance legal responses to citizen data not only as a subject of potential violation, but also as a potential resource for citizen empowerment. Dignitarian governance systems like the GDPR may establish standards of violation and pathways for exit, but these protodemocratic forms of public-data governance offer a promising (and largely, though not always complementary) addition to grow and develop public capacity to utilize data infrastructure for public ends.[178] In other words, rather than a governance approach that establishes what private entities may *not* do to German, Canadian, or Barceloní citizens' data, these alternative approaches consider what data as a public resource *can* do for German, Canadian, or Barceloní citizens. Indeed, the

176. Dame Wendy Hall & Jerome Pesenti, *Growing the Artificial Intelligence Industry in the UK*, GOV.UK (Oct. 15, 2017), https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk [https://perma.cc/8CT3-8RB4] (recommending data trusts to improve secure and mutually beneficial data exchanges). Ontario has commissioned a series of discussion papers for the region's Data Strategy, which includes discussion of the merits of data trusts, and also launched a public consultation session in August 2020 to seek public input. *See Ontario Launches Consultations to Strengthen Privacy Protections of Personal Data*, ONTARIO (Aug. 13, 2020), https://news.ontario.ca/en/release/57985/ontario-launches-consultations-to-strengthen-privacy-protections-of-personal-data [https://perma.cc/FHE9-5KPP]. The Open Data Institute is a prominent international nonprofit group that works with governments and other entities to develop more open data ecosystems and has worked with the UK government (among others) to research and implement data trusts. *See Data Trusts: Lessons from Three Pilots*, OPEN DATA INST. (Apr. 15, 2019), https://theodi.org/article/odi-data-trusts-report [https://perma.cc/9MEZ-SMGA]. For more on data trusts generally, see Bianca Wylie & Sean McDonald, *What Is a Data Trust?*, CIGI ONLINE (Oct. 9, 2018), https://www.cigionline.org/articles/what-data-trust [https://perma.cc/LBZ5-WANV].

177. EVGENY MOROZOV & FRANCESCA BRIA, RETHINKING THE SMART CITY: DEMOCRATIZING URBAN TECHNOLOGY 27-29 (Jan. 2018), https://rosalux.nyc/wp-content/uploads/2021/02/RLS-NYC_smart_cities_EN.pdf [https://perma.cc/Q99E-FKSA] (detailing Barcelona's approach to building a "city data commons").

178. The dignitarian data-subject rights granted under the GDPR may provide a complementary backstop to the kinds of affirmative data production envisioned by such proposals, but as discussed in the Waterorg example, strong individual data-subject rights may also foreclose them. In fact, many commentators believe the proposed Data Governance Act in the EU, which provides the basis for some collective forms of data governance, would violate fundamental data-subject rights in the EU, because it would allow data subjects to devolve inalienable rights over their data to the data institutions.

highly attuned feedback structures that data production allows offer new possibilities for public governance and social coordination.

Not all proposals advocating for new collective-data institutions envision traditionally public forms of data management. Others seek to democratize governance of data production as part of ongoing efforts to democratize other spheres of life, most notably the workplace. Labor activists are developing worker-data collectives to counter growing workplace surveillance by employers by monitoring forms of workplace oppression and documenting Occupational Safety and Health Act (OSHA) violations and wage theft, with the goal of collectively negotiating how algorithms govern life at work.[179] Other advocates are developing alternative worker-based data streams to better document the economic value and impact of essential workers, or to give workers greater ability to document and trace supply chains for their products.[180] These nongovernmental collective alternatives may be particularly attractive in places and with respect to data flows where individuals have little faith either in private companies or the government to safeguard collective interests.[181] Private data-governance mechanisms may also face certain challenges in realizing the ideals of democratic-data governance. Most notably, proposals for private trusts generally

179. WeClock, https://www.weclock.it/about [https://perma.cc/UBR7-TD5G] ("[WeClock] offers a privacy-preserving way to empower workers and unions in their battle for decent work."); *Lighthouse: A Guide to Good Data Stewardship for Trade Unions*, Prospect, https://lighthouse.prospect.org.uk [https://perma.cc/63LY-D8X9]. The National Domestic Workers' Alliance developed its alternative platform for domestic workers to help house cleaners get benefits by providing clients a platform to contribute to a cleaner's Alia count. In turn, cleaners can use the collective contributions from clients to purchase benefits that domestic workers may not otherwise be entitled to by law. *See Alia*, Nat'l Domestic Workers' All., https://www.ndwalabs.org/alia [https://perma.cc/Z76C-9384].

180. *La Alianza*, Nat'l Domestic Workers' All., https://www.ndwalabs.org/la-alianza [https://perma.cc/Q8PB-38FX]; *see* Katya Abazajian, *What Helps? Understanding the Needs and the Ecosystem for Support*, Mozilla Insights 37-38 (Mar. 31, 2021), https://foundation.mozilla.org/en/data-futures-lab/data-for-empowerment/what-helps-the-ecosystem-for-needs-and-support [https://perma.cc/B2DZ-BKTD]. For example, Abalobi gives South African fishing communities access to data that helps them track where their fish is sold and connect with restaurants and other patrons who buy their stock; the platform is managed by fishing-labor cooperatives that make collective decisions regarding the platform. Abazajian, *supra*.

181. In an international survey of several organizations developing alternative data-governance regimes conducted by Mozilla, almost all respondents suggest that users would trust a collective of peers more than they would trust themselves or government to appropriately use their data. *See* Abazajian, *supra* note 180, at 36.

work by pooling individual data-subject rights.[182] These are often designed to only recognize the interests of data subjects from whom data is collected, rather than also considering those on whom data products may be used—and who therefore also has a relevant interest in the terms that govern how data is collected and processed.[183]

Finally, existing forms of trusted public-data collection and management, like those of the U.S. Census and its statistical agencies, the Library of Congress, and state and local municipal libraries, may be expanded into more general data-governance bodies.[184] Public statistical agencies and libraries have established professional expertise around responsible information and knowledge management for the public good, and adhere to strict purpose limitations as well as high confidentiality standards.[185] Alternatively, public-data management for the public good may be achieved via an expanded remit for scientific research agencies such as the National Institutes of Health and the National Science Foundation,

182. *Enabling Data Sharing for Social Benefit Through Data Trusts*, GLOB. P'SHIP ON A.I., https://gpai.ai/projects/data-governance/data-trusts [https://perma.cc/4ATP-AZ4G] (defining data trusts as "a form of data stewardship that allow data producers to pool their data (or data rights) and facilitate collective negotiation of terms of use with potential data users, working through independent trustees who are bound by strong fiduciary duties, within a framework of technical, legal and policy interventions that facilitate data use and provide strong safeguards against mis-use"); *see also Data Trusts: A New Tool for Data Governance,* ELEMENT AI & NESTA, https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf [https://perma.cc/TL82-BV9E] (advancing the public-policy conversation on data trusts and the need for improved data governance); OPEN DATA INST., *supra* note 176 (exploring ways of increasing access to data while retaining trust).

183. In response to growing civic activism against their proposed Waterfront "smart" neighborhood in Toronto, Sidewalk Labs (an urban innovation subsidiary of Alphabet, Inc.) proposed a civic-data trust to manage the urban data collected in the neighborhood. The proposal conflated the social interests that arise from collection and use in a number of ways, most notably by bundling the licenses for use and collection and by increasing proprietary control over more sensitive (and more valuable) personal information. *See* Sean McDonald, *Toronto, Civic Data, and Trust*, MEDIUM (Oct. 17, 2018), https://medium.com/@digitalpublic/toronto-civic-data-and-trust-ee7ab928fb68 [https://perma.cc/XUR4-QWQR].

184. Jake Goldenfein, Ben Green & Salomé Viljoen, *Privacy Versus Health Is a False Trade-Off*, JACOBIN (Apr. 17, 2020), https://jacobinmag.com/2020/04/privacy-health-surveillance-coronavirus-pandemic-technology [https://perma.cc/VG6W-YBUM]; JULIA LANE, DEMOCRATIZING OUR DATA (2020); Sadowski et al., *supra* note 173.

185. U.S. Census Act, 13 U.S.C. §§ 8(b)-(c), 9 (2018); Eun Seo Jo & Timnit Gebru, *Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning*, FAIRNESS, ACCOUNTABILITY & TRANSPARENCY (Jan. 2020), https://dl.acm.org/doi/pdf/10.1145/3351095.3372829 [https://perma.cc/8QQN-X366].

or public agencies that already hold public data like the Food and Drug Administration.[186] These agencies already have institutional expertise in stewarding data and managing scientific resources in service of the public good. While none are perfect, each stems from long professional histories of managing collective knowledge in the public interest.

b. The Possibility of Democratic Data

The data economy has resulted in massive collection of information regarding consumer-purchasing preferences and social networks, but it has contributed comparatively little to ongoing discussions concerning waste production, water usage, or how wealth from financial instruments flows globally.[187] Companies know a great deal about their consumers, but consumers still have little insight into the supply chains, ownership structures, and operating practices of companies. Workers are subject to increased surveillance at the workplace and in the screening process for employment, but know comparatively little about the hiring practices, quality of workplace life, and histories of discrimination and harassment of employers. Ensuring greater recognition can expand the set of interests considered relevant to setting the agendas of data production, and in turn how data infrastructures are funded and developed. In short, conceiving of data's democratic possibilities can provide greater standing for a wider range of priorities and goals to motivate how and why information is produced. This may result not just in less consumer-preference data production, but also in the proliferation of other kinds of socially useful data production.

As the Waterorg example shows, DDM also affords stronger conceptual footing for data-production conditions that may require 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. This has important implications for other public-reform projects that will

---

186. *See* Christopher J. Morten & Amy Kapczynski, *The Big Data Regulator, Rebooted: Why and How the FDA Can and Should Disclose Confidential Data on Prescription Drugs and Vaccines*, 109 CALIF. L. REV. 493, 493 (2021).

187. *Compare* Einav & Levin, *supra* note 20, at 718 ("Private companies that specialize in data aggregation, such as credit bureaus or marketing companies such as Acxiom, are assembling rich individual-level data on virtually every household"), *with* Richard Henderson & Owen Walker, *BlackRock's Black Box: The Technology Hub of Modern Finance*, FIN. TIMES (Feb. 24, 2020), https://www.ft.com/content/5ba6f40e-4e4d-11ea-95a0-43d18ec715f5 [https: //perma.cc/S79A-FA8W] (discussing how BlackRock's tech platform Aladdin acts as a "central nervous system for many of the largest players in the investment management industry"). BlackRock is not required to disclose how many of the world's assets sit on the system. They last did so in 2017, at which time they reported $20 trillion; since then, BlackRock has added scores of new clients.

almost certainly rely upon data infrastructures and citizen data. Conceptually distinguishing and defending data production for core public functions are especially valuable for data-governance reform projects that act from a political position that citizens are owed more, not less, from the state by virtue of their status as citizens.[188] Public provisioning will require making productive and distributive decisions over social resources — decisions that should be (and indeed, likely must be) informed by citizen data. The data infrastructures necessary to responsibly produce and allocate goods and services, such as healthcare, education, housing, clean air, and fresh water, will require some degree of mandatory citizen-data collection to manage this provision efficiently and fairly.

3. Democratic Regimes and Individual Data-Subject Rights

The discussion above highlights a few key insights regarding the relationship between legal agendas for democratic-data governance and those that prioritize individualized data-subject rights.

First, the theory of democratic regimes advocated in this Feature is agnostic regarding the ontological commitments implied by individualist regimes that view data either as "thing-like" or "person-like." There is a long philosophical (and legal) tradition that makes sense of both property and persons as constitutive of and constituted by social relations. Where democratic-governance proposals depart from individualist ones is in their conception of where interests in information adhere, and the legal agendas that flow from this conception.

For instance, democratic-governance regimes clearly do not conceptually negate the notion that data is being treated as an asset, or that individuals have an interest in how such assets are produced and used to create value (and social harm). They repudiate the idea that we can reduce the social interests we have in data-value production and its distribution to individual propertarian interests in social-data resources. If data assets are to be viewed as a kind of property, a DDM account supports attempts to govern it via public trusts or similar kinds of common-ownership institutional arrangements, and challenges attempts to distill individual legal claims to value from collective-data value. It also challenges the notion that the social interests people have in data value are purely indexed in distributions of monetary value: distributing the spoils of exploitative data relations does not equalize those data relations.

---

188. Elizabeth Anderson, *How Should Egalitarians Cope with Market Risks?*, 9 THEORETICAL INQUIRIES L. 239 (2008); *see also* Elizabeth Anderson, *Common Property: How Social Insurance Became Confused with Socialism*, BOS. REV. (July 25, 2016), http://bostonreview.net/editors-picks-us-books-ideas/elizabeth-anderson-common-property [https://perma.cc/5AXG-UXR6] (describing different forms of social insurance that states provide their citizens).

Second, let us consider the stronger (and more legally well-developed) challenge posed by the dignitarian conception of data-subject rights. Democratic-governance regimes do not repudiate the notion that individuals have dignitary interests in information. It repudiates the idea that legal protection of these interests is reducible to the vertical relation between data subject and data collector.[189] Consider, for example, data collected by a fertility-tracking app suggesting that a person (let's call her Amy) is in her first trimester of pregnancy. One may consider it a dignitary violation for an advertising company or employer to gain downstream access to this data about Amy. But Amy's dignitary interests in keeping her pregnancy private are implicated whether the company gains access to Amy's data via her fertility-tracking app, or whether the company contracts with a service that analyzes and infers from several relevant features Amy shares with known pregnant people that there is a ninety-five percent chance that Amy is in her first trimester of pregnancy. Amy has a dignitarian interest against people seeking to learn her pregnancy status, but this interest resides — both for Amy and for others like her — at the category level of first-trimester pregnancy data.

Democratic regimes also allow us to recognize (and adjudicate among) competing dignitarian interests with respect to the same data. For instance, responding to Amy's dignitarian interests by restricting the collection of first-trimester pregnancy data may be in tension with the dignitarian interests of others to enact their informational self-determination — to share data about their first trimester pregnancy with a fertility app to enjoy its services.

But as this Feature has endeavored to show, people do not only have dignitarian interests in information; they also have egalitarian ones. These interests index concerns over social informational harm: that people have a collective interest against the unjust social processes data flows may materialize, against being drafted into the project of one another's oppression as a condition of digital life, and against being put into data relations that constitute instances of domination and oppression for themselves or others on the basis of group membership. Casting all relevant concerns regarding information as individual claims to payment or self-determination masks collective egalitarian social interests in enacting data relations of equality (and addressing data relations of oppression).

By recognizing such interests, democratic-data regimes in turn apprehend potential tensions (both conceptual and institutional) between achieving more egalitarian-data relations and robust dignitarian informational protections. Consider again Amy's first-trimester pregnancy data. The dignitarian account

189. It is important to note that even under an expansive democratic regime, certain dignitarian interests in information *do* rightly reside with individual data subjects. For instance, democratic data regimes should grant individuals rights against being singled out or reidentified by aggregate data processing that is meant to provide insight into population-level trends, as well as rights over unique biometric identifiers for purposes of identification and verification.

may well express that companies or employers gaining access to this information would violate a privileged relationship Amy enjoys to this sensitive information. But a relational account of this data flow also captures why this information is so sensitive to begin with.

One may find Amy's data flow particularly sensitive because of its significance in constituting a relevant group identity — of materializing a key aspect of what it means (legally and socially) to occupy the status of "woman" in this particular historical context.[190] Part of the social construction of womanhood involves the contested legal and social terrain of early pregnancy.[191] Data flows that impart knowledge of an early pregnancy bring Amy onto this terrain. This in turn leaves her vulnerable to certain forms of social oppression on the basis of this category membership. It may implicate or constrain the choices she makes (including sensitive and contested ones like terminating her pregnancy) that are intimately bound up with how legal, cultural, and social institutions construct and condition womanhood. In sum, early pregnancy data flows are sensitive and require governance because these flows help to materialize social relations of sex, gender, and fertility — and depending on how these data flows are governed, they can exacerbate or reduce the inegalitarian condition of these relations.

Put differently, many of the intuitions regarding data flows currently cast in the language of dignitarian interests actually have a great deal to do with the capacity of data flows to materialize salient social relations. For instance, pregnancy data is sensitive because "pregnant women" is a historically oppressed social category. The data flow "redheads who like cats" likely implicates far fewer (and far less significant) legal interests because "redheads who like cats" is not a social category historically constituted through domination.

But to distinguish between data flows that constitute socially innocuous categories and socially consequential ones, and to distinguish between (and adjudicate among) social egalitarian interests and individual dignitarian ones, requires comprehensive data-governance mechanisms that can apprehend these various interests at the population level.

Democratic-governance regimes depart from individualist alternatives in recognizing the plurality of (population-level) interests in information production, and in providing a normative theory for adjudicating among them. The

190. For the sake of this argument, we will assume that Amy identifies as a woman. But of course, people who do not identify as women can also become pregnant. The social relation materialized by this data flow may be considered even more sensitive and more significant in the case where Amy does not identify as a woman.

191. Here too, the social condition of early pregnancy may not only typify and demarcate the social meaning of womanhood but may also describe the condition of people who are pregnant and do not identify as women. In such instances, the condition of early pregnancy is likely even more legally and socially contested.

underlying claims of injustice that motivate individualist agendas for reform are important but incomplete. Reducing these interests to individual data-subject rights in a data transaction gives short shrift to these interests, and fails to apprehend when other interests may conflict with and at times supersede such interests.

CONCLUSION: REORIENTING THE TASK OF DATA GOVERNANCE

If the aim of data governance is to account for population-level interests in the digital economy, then different legal conceptions of informational harm (and our legal responses to them) may be required. This is not to say that injustice may not also occur along vertical relations—it may, and it does. But, as Part II establishes, the imperatives to relate people to one another place pressure on the conditions of exchange that structure vertical relations; accounting for population-level horizontal interests is thus relevant to the task of addressing these forms of injustice, too.

Then, Part III shows that theories of data governance that stem from individualist conceptions of informational harm do not represent the social effects of data production as a result of the pervasive population-level horizonal relations that data production enacts. Such theories thus cannot address the ways these effects may cause harm or how these effects could be structured to produce shared benefits. This presents a methodological limitation and an epistemic deficiency, since such notions of informational harm fail to provide adequate tools for identifying and addressing the harmful social effects that datafication produces.

The conceptual account offered by this Feature foregrounds data's relationality, which results in a few helpful reorientations regarding the task of data governance. First, it clarifies that social inequality is not a byproduct of unjust data collection, but is an injustice of concern in data production in its own right. This informs a different diagnosis of data-governance failure. On this account, datafication may be wrong not only because it manipulates people, but also because the social effects it produces or materializes violate standards of equality. As an economic process, datafication may lead to unfair wealth inequality that violates distributive ideals of justice. As a social process, datafication may reproduce and amplify forms of social hierarchy that violate relational standards of justice.

The prevalence of population-level interests in data production means that one's actions in the data political economy necessarily impact others in uneven ways over which one has no direct control, often recreating or exacerbating the

durable inequalities that operate along the lines of group identity.[192] This raises quintessentially democratic questions: it requires negotiating trade-offs among groups of people with competing and at times normatively distinct interests. Hence, datafication gives rise not only to personal claims regarding risk of personal violation that justify personal ordering, but also to population-level claims about the risk of social effects that justify political ordering.

The unsettled status of data in law presents both a challenge and an opportunity: a challenge for addressing the injustices that arise from digital life, and an opportunity to experiment with the kinds of social ordering the law may enact in response. Far from offering terrain on which to reimpose forms of private market ordering or narrow civil-libertarian claims, data governance may plausibly retrieve spheres of life from private governance and begin to develop new alternatives.

192. TILLY, *supra* note 170; YOUNG, *supra* note 159. Nick Couldry and Ulises A. Mejias offer one interesting account theorizing the data-social relation as that of colonizer and colonized, an account they refer to as "data colonialism." Couldry & Mejias, *supra* note 85.