

Data Unions: The Need for Informational Democracy

Eli Freedman*

The data that everyday consumers produce is becoming more and more important to the economy. Yet, as this data imbues tech corporations with tremendous wealth and power, we, the data producers, have no say as to how our data is collected or how it is used. The reign of data analytics to pursue profit above all else has led to a conflagration of data harms perpetuated against already marginalized groups. What is needed in this moment is a tool that equalizes the bargaining power between platforms and users, to give consumers meaningful control over the data they produce. In the early 20th century, labor organizers called for industrial democracy: the ability for workers to have substantial say over the conditions of their labor. For today's datafied information economy, this Note instead calls for the need for informational democracy: the ability of consumers, as data producers, to exert meaningful control over the data that their lives engender.

This Note advocates for data unions as one such tool to achieve informational democracy. It conceptualizes data unions as democratically elected organizations that aggregate data to create collective bargaining units to negotiate with platforms as to allowed uses for data. First, the Note gives an overview of how today's economy creates both value and harm out of data processing. Then, it argues that due to the specific nature of this value and harm creation, data unions are uniquely situated regulatory tools that can enact meaningful consumer control.

DOI: <https://doi.org/10.15779/Z38WS8HN0X>

Copyright © 2023 Eli Freedman.

* J.D., 2022, University of California, Berkeley, School of Law. I am very grateful to Professor David Singh Grewal for guiding and supervising this paper. I am thankful to Imienfan Uhumwuangho for her willingness to lend an extra set of eyes and thoughtful critiques. Many thanks to the editors of the *California Law Review* for their patience and innumerable contributions.

Introduction	658
I. The Political Economy of Data is Predicated on Aggregate Data, Not on Individual Data	662
A. Data’s Value Is Driven by Horizontal Predictions Between Data Subjects	662
B. The Data Political Economy Also Produces Harms Based on Horizontal Relations	666
C. The Current Data Regulatory Structure Is Ineffective Against Both Vertical and Horizontal Harms	667
D. Proposed Data Reforms Also Fail to Address Horizontal Harms.....	669
II. The Need for Informational Democracy	671
A. Unions Increase Wages and Protect Worker Dignity	673
B. For the Data Unions Make Us Strong.....	674
C. Data Unions Could Protect Against Vertical and Horizontal Harms.....	677
Conclusion	680

INTRODUCTION

In our hands is placed a power greater than their hoarded gold[,] Greater than the might of armies[,] magnified a thousandfold[,] We can bring to birth a new world from the ashes of the old For the Union makes us strong.¹

—Early 20th Century Trade Union Anthem

Information machines are the sole means of vision in digital visual culture, but as the body itself becomes socially defined and handled as information, there is even more at stake in paying attention to the incursion of machines in everyday life and the forms of resistance available to us.²

—Lisa Nakamura

There are now two of each of us. One is familiar to us and under our control. The other is shrouded in secrecy and jealously guarded beyond our control, by powerful private corporations. The “familiar you” is just you, flesh and firing brain synapses, reading this Note. The “other you” is your data double, constructed by platforms from the digital trails you have left behind to be the

1. Ralph Chaplin, *Solidarity Forever*, GENIUS, https://genius.com/Ralph-chaplin-solidarity-forever-lyrics?__cf_chl_captcha_tk__=Kz6iArJCO8BPGaZnMgkh7C5szfmRdoayvCtCzM8cA3Q-1639942467-0-gaNycGzNCf0 [https://perma.cc/TN7K-X7P3].

2. LISA NAKAMURA, DIGITIZING RACE: VISUAL CULTURES OF THE INTERNET 130 (2008).

quantified embodiment of your features, likes, and predictability.³ Having access to our data doubles enables tech corporations to make population-wide predictions from data analytics. It is this process, the creation of a digital replica of your online likes and clicks, that drives the incredible profits and power of the new data economy.⁴

From smart toilets to smart refrigerators to smart electronic plugs, the minutia of daily life is becoming traceable in data. And this data goes beyond just what utensils or social media posts we like. In January 2022, it came to light that one of world's most popular suicide and mental health support lines, Crisis Text Line, was using and monetizing its trove of mental health data to market customer service software.⁵ In the context of people and communities existing and moving through online and technological spaces, data is nothing more than the commodification of our lives.⁶ As the Crisis Text Line example makes clear, even our most private and intimate moments are commodifiable.⁷ In today's economy, the data our lives produce is becoming increasingly "essential" for a wide array of economic sectors, like "technology, infrastructure, finance, manufacturing, insurance, and energy."⁸ For example, Crisis Text Line used consumer data to teach tech industry artificial intelligence (AI) chatbots how to live chat with customers seeking redress or help.⁹ As the value and need for data inputs increases, so too does the corporate drive to extract more and new kinds of data from us.¹⁰

The data derived from our simple acts of living, of doing, or of moving through the world has created tremendous value for the companies that collect, refine, analyze, and sell our data.¹¹ It has also created tremendous private power, with corporations able to sculpt the world around us by choosing what information we see¹² and what products we can buy.¹³ Like a positive feedback loop locked in overdrive, our engagement with these platforms creates more data

3. See JULIE E. COHEN, *BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM* 67 (2019).

4. See *id.*

5. Alexandra S. Levine, *Suicide Hotline Shares Data with For-Profit Spinoff, Raising Ethical Questions*, POLITICO (Jan. 28, 2022), <https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617> [<https://perma.cc/2ESY-T67H>].

6. See NICK COULDRY & ULISES A. MEJIAS, *THE COSTS OF CONNECTION: HOW DATA IS COLONIZING HUMAN LIFE AND APPROPRIATING IT FOR CAPITALISM* 6–7 (2019).

7. See Levine, *supra* note 5.

8. Jathan Sadowski, *When Data Is Capital: Datafication, Accumulation, and Extraction*, 6 *BIG DATA & SOCIETY* 1, 1 (2019).

9. Levine, *supra* note 5.

10. See Sadowski, *supra* note 8, at 4.

11. See *id.*

12. FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* 130 (2015).

13. See Adrienne Jeffries & Leon Yin, *Amazon Puts Its Own "Brands" First Above Better-Rated Products*, MARKUP (Oct. 14, 2021), <https://themarkup.org/amazons-advantage/2021/10/14/amazon-puts-its-own-brands-first-above-better-rated-products> [<https://perma.cc/TB22-CB5H>].

about our preferences, resulting in more value and power for these corporations. And so, the process whirls on, faster and faster.

Yet even though our lives, crafted into data inputs, are the engine driving the new political economy for datafied information, we individually or collectively have no meaningful control over our data. Corporations acquire our data through long boilerplate contracts that leave no room for us to bargain meaningfully.¹⁴ In other situations, our data is collected without any consent as corporations secretly track our footsteps across the web.¹⁵ Individually, the data we produce is of the slimmest economic value, small enough to be incalculable.¹⁶ However, collectively, the value of our data is immense.¹⁷

In the early 20th century, when workers suffered from little control over their labor conditions, organizers called for industrial democracy—that is, greater work control over the circumstances of their work.¹⁸ This Note looks to this history for inspiration. The rallying around industrial democracy embodied workers' demand to have a greater say and control over their workplaces, and by extension, their lives.¹⁹ Inspired by the rallying cry for greater laborer control of their daily work lives, this Note calls for *informational democracy*: the ability of people, currently turned into commodified data subjects for capital, to meaningfully control the data they produce. Our data doubles as a predictive tool that directly impacts our daily lives, from our credit scores, to our mortgage access, to even our entitlement to social welfare from the state.²⁰ To remedy the issues of absolute control by tech corporations and no meaningful agency on behalf of data producers, this Note envisions a solution in data unions: democratically controlled aggregated data pools that allow for individuals to meaningfully control their data as a collective bargaining unit. In the unionization of data, this Note aims to strike a path towards genuine informational democracy and consumer control over data.

14. See Sadowski, *supra* note 8, at 7–8.

15. See David Ingram, *Facebook Fuels Broad Privacy Debate by Tracking Non-Users*, REUTERS (Apr. 15, 2018), <https://www.reuters.com/article/us-facebook-privacy-tracking-idUSKBN1HM0DR> [<https://perma.cc/CH53-PPRN>]; Dr_Jeff, *They Store Data on You Even If You Did Not Interact with the Service*, TERMS OF SERV. DIDN'T READ, <https://edit.tosdr.org/points/12804> [<https://perma.cc/XVR2-HDCE>]; Angie Waller & Colin Lecher, *Help Us Investigate Facebook Pixel Tracking*, MARKUP (Jan. 21, 2022), <https://themarkup.org/news/2022/01/21/help-us-investigate-facebook-pixel-tracking> [<https://perma.cc/Y74K-C8AZ>]; Lily Hay Newman, *Health Sites Let Ads Track Visitors Without Telling Them*, WIRED (Feb. 6, 2022), <https://www.wired.com/story/health-site-ad-tracking> [<https://perma.cc/2HYN-BDWE>].

16. See Sadowski, *supra* note 8, at 8.

17. See *id.*

18. JOSEPH A. MCCARTIN, LABOR'S GREAT WAR: THE STRUGGLE FOR INDUSTRIAL DEMOCRACY AND THE ORIGINS OF MODERN AMERICAN LABOR RELATIONS, 1912–21 12–13 (1997).

19. See *id.*

20. See COULDRY & MEJIAS, *supra* note 6, at 131; Karen Hao, *The Coming War on the Hidden Algorithms That Trap People in Poverty*, MIT TECH. REV. (Dec. 4, 2020), <https://www.technologyreview.com/2020/12/04/1013068/algorithms-create-a-poverty-trap-lawyers-fight-back/> [<https://perma.cc/N95L-E4XF>].

This Note adds two critical elements to the discussion surrounding data regulatory regimes and data unions. Others have written about combining data into union pools. However, these other works only envision unions as tools to monetize data production for everyday users. Phrased another way, the purpose of these data unions is to ensure that people receive compensation for the data they produce.²¹ This Note does not conceptualize data unions as a method for users to receive a data production wage. This is because a central concern of this Note, as will be argued later, is the way that data collection and algorithmic ordering threaten human autonomy.²² If data unions were just tools for monetization, they would not deal with autonomy concerns. Instead, they could have the opposite effect because paying people for data could incentivize people to hand even more data in a way that might undermine human autonomy. This Note's conception of data unions is not about monetization but is about fixing the sheer power imbalance between platforms and users to ensure that people have meaningful control over their data.

The second distinction is that this Note takes a novel approach to data harms. In her article, Salomé Viljoen argued that current data regulatory regimes and proposals do not properly deal with data harms because they misunderstand the foundational factor, horizontal relations between data subjects, that drives the value of data in today's political economy of informational capitalism.²³ In turn, this Note builds on Viljoen's incisive critique to show why data unions, as opposed to other proposed regulatory structures, adequately deal with how the data economy produces value for corporations while harming users.

This Note proceeds in the following way. Part I makes clear what the structure of the political economy surrounding data is. It further explains how value is created from data and how these structures harm individuals, particularly those from discriminated against communities. It then shows how current and proposed data regulations fail to accurately deal with the structure of the data economy. Part II argues that data unions would be an efficient regulatory device for the issues described in Part I, and it imagines how data unions could look and function.

21. See THE DATA UNION, <https://www.thedataunion.org/> [<https://perma.cc/WD8R-37XJ>]; ERIC A. POSNER & E. GLEN WEYL, *RADICAL MARKETS: UPROOTING CAPITALISM AND DEMOCRACY FOR A JUST SOCIETY* 205–07, 209, 242–43 (2019); Tom Hamilton, *What Are Data Unions? How Do They Work? Which Ones Can I Use?*, MEDIUM (Mar. 31, 2020), <https://medium.com/streamrblog/what-are-data-unions-how-do-they-work-which-ones-can-i-use-887e67fb7716> [<https://perma.cc/3PRC-2K7U>].

22. See, e.g., COULDRY & MEJIAS, *supra* note 6, at 156–57.

23. See Salomé Viljoen, *Democratic Data: A Relational Theory for Data Governance*, 131 *YALE L.J.* 573, 582 (2021).

I.

THE POLITICAL ECONOMY OF DATA IS PREDICATED ON AGGREGATE DATA,
NOT ON INDIVIDUAL DATA

To understand why data unions would be effective regulatory mechanisms for the datafied economy, it is first necessary to understand how value is created out of data and how data practices harm individuals. In her important article, *Democratic Data*, Salomé Viljoen critically argued that the data economy must be understood as driven by the horizontal relationships between data subjects (the people who data is collected from).²⁴ Viljoen's view is important because it differs from the majority of work on data regulations that focuses only on the relationship between the individual and the platform collecting the data.²⁵ This Note heavily relies on Viljoen's argument to show why data unions are an appropriate regulatory fit for the datafied economy. In this Section, the Note will show how data's value and harms are driven by horizontal relationships between data subjects and how this harm disproportionately affects marginalized groups. It will then note how current and proposed regulations for data governance fail because they do not account for the horizontal nature of data's political economy.

A. *Data's Value Is Driven by Horizontal Predictions Between Data Subjects*

Data's value is not predicated on our individually quantified metrics, but on what our data, taken together with other similarly situated people, reveals about population-wide trends. Phrased differently, our data's value is not based on the vertical relationship between us and the data collector, but on the horizontal relationship between us and other data subjects and what it reveals about specific demographic groups.²⁶ A vertical data social relation is the movement of data from data subject to data collector, like from a Facebook user to Facebook.²⁷ A legal structure that only regulated vertical relationships would only be concerned about the power dynamic between users and the platforms they use. For example, these laws would target what data Facebook or Twitter could get from consumers. The horizontal data relationship is not between an individual data subject and collector, but between multiple data subjects, indicating population-wide conclusions.²⁸ A legal structure targeting horizontal relationships would heed the way that data aggregated into specific demographic groupings, like age, could predicate trends among such a group. It is horizontal because data about one set of twenty-year-old data subjects horizontally impacts other twenty-year-olds, because the corporation is now able to make grouping-wide predictions. An example of a horizontal relationship would be the

24. *See id.* at 609.

25. *See id.* at 603–13.

26. *See id.* at 609.

27. *See id.* at 607.

28. *Id.*

aggregation of a segment of similar Facebook users by geographic location, age, or another quantifier, and what that data reveals about their preferences.

Before this Note can argue why unionization for data works as a democratic data governance technique, it needs to first make clear how the political economy of informational capitalism assigns data value in the first place. Legal scholar Julie Cohen named the new era of capitalism that we are now a part of as “informational capitalism.”²⁹ This term refers to the “alignment of capitalism as a mode of production with informationalism as a mode of development.”³⁰ She clarified that as capitalism “is oriented toward profit-maximizing, that is, toward increasing the amount of surplus appropriated by capital on the basis of private control over the means of production and circulation,” while informationalism “is oriented . . . toward the accumulation of knowledge and towards higher levels of complexity in information processing.”³¹ Therefore, in our current era of informational capitalism, “market actors use knowledge, culture, and networked information technologies as a means of extracting and appropriating surplus value.”³²

In informational capitalism, a new commodity is of critical importance: data. Data can be understood as the commodification of human life,³³ where the traces of what we do, what we buy, who we interact with, where we go, etc., become inputs to create surplus value for extraction. To facilitate access to data, informational capitalism is dictated by an ideology of datafication, “which insists that every aspect of life must be transmuted into data as the form in which all life becomes useful for capital.”³⁴ Yet to achieve datafication, corporations must have a way to control the flow of extracted data. To accomplish this, informational capitalism encloses and creates a corporate semi-property interest of the formerly intangible resource of data.³⁵

While data is not formally recognized as a kind of intellectual property, overlaps in contract law and trade secret law have effectively rendered data as a *de facto* form of property that users have no control over.³⁶ Phrased differently, despite data not being treated as *de jure* intellectual property, corporations have effectively exploited contract and trade secret law to create property-like privileges over the data they collect.³⁷ Platforms use these two areas of law to create zones that allow them to exclude any other party from reaching the company’s own data trove.³⁸ To access internet platforms, users are required to

29. COHEN, *supra* note 3, at 5.

30. *Id.*

31. *Id.* at 5–6.

32. *Id.* at 6.

33. See COULDRY & MEJIAS, *supra* note 6, at 6–7.

34. *Id.* at 16.

35. See COHEN, *supra* note 3, at 15.

36. See *id.* at 44.

37. See *id.*

38. See *id.*

agree to boilerplate contracts.³⁹ These contracts are often the legal starting point of the vertical data relationship.⁴⁰ At our point of access to the service, platforms require that users give full control over data, while voiding the ability of users, third-party vendors, and advertisers to understand the platform's algorithm and data collection processes.⁴¹ Therefore, at the initial point of contact with a platform, users lose any control over the data they produce. Rather than a contract of meaningful assent, these boilerplate contracts are nonnegotiable. People are likely to click "Yes" not because they understand the terms, but because they acknowledge that consenting to the agreement is the only path to the service.⁴²

To make matters worse, corporations often acquire our data even without the modicum of consent embodied in boilerplate contracts as discussed above. Both Twitter and Facebook collect data on people who do not have accounts, yet end up on a webpage that has a link to a Twitter or Facebook page or like button.⁴³ Facebook maintains and updates data profiles on users who have deactivated their accounts, which raises questions as to whether users can ever meaningfully opt out of Facebook's surveillance after initially signing up for an account.⁴⁴ Furthermore, Facebook embeds code in websites that allows them to track users without consent.⁴⁵ It is currently unclear how extensive the reach of Facebook's tracking across the web is.⁴⁶ Everyday internet users have no say over how their data is collected, and therefore no control over the data's subsequent uses.

Once platforms have acquired users' data, the corporations must refine it. The acquired data flows are "processed to generate patterns and predictions about data subjects' preferences and behaviors."⁴⁷ Platforms create data doubles of all users from these generated patterns and predictions. Each data double is tied to an actual person.⁴⁸ After collecting users' activities, likes, preferences, connections, etc. in data form, the data double becomes the refined embodiment of our simulated behavior.⁴⁹ The data double's purpose is to "make human behaviors and revealed preferences calculable, predictable, and profitable *in aggregate* . . . they are designed to enable the statistical construction,

39. See Viljoen, *supra* note 23, at 598, 607.

40. See *id.* at 594, 598.

41. See *id.*

42. See Sadowski, *supra* note 8, at 8–9.

43. Ingram, *supra* note 14; Dr_Jeff, *supra* note 15.

44. See Kate Kaye, *Why Facebook Keeps Collecting People's Data and Building Their Profiles Even When Their Accounts Are Deactivated*, DIGIDAY (Oct. 28, 2021), <https://digiday.com/media/why-facebook-keeps-collecting-peoples-data-and-building-their-profiles-even-when-their-accounts-are-deactivated/> [https://perma.cc/M5S7-52B4].

45. Newman, *supra* note 15.

46. See WALLER & LECHER, *supra* note 15.

47. COHEN, *supra* note 3, at 66.

48. *Id.* at 67.

49. See *id.*

management of, and trade in populations.”⁵⁰ Aggregation of data doubles into specific demographic segments, like race, age, sexual orientation, and religion, creates value. Each segmented group then has “probabilistically determined behavioral profiles” that companies buy to reach consumers likely to want their product.⁵¹ This is why, as noted before, regulation should focus on the horizontal data relationship between two data subjects. The utility of the data doubles is in pooling them together so that specific demographic-wide predictions are quantifiable. For example, in combining the data doubles of those from a like-income, like-age, like-geographic-area grouping, corporations can target specific trends about that group’s likes, dislikes, tastes, and shopping habits. Phrased differently, data’s value is not in tracing the wants of a specific individual, but in the ability to predict, based on aggregated sets of data doubles, what a targeted demographic group will want or care about.

Data cultivators derive value from data by selling their crafted data double tranches into data markets.⁵² Data markets are markets where businesses and other organizations can purchase data double pools as inputs into their own production processes.⁵³ Again, purchasers are not buying a single data double itself, but rather specific demographic groupings of data doubles that have been constructed for a predetermined purpose: predictions on this group’s likes and preferences.⁵⁴ It is important to highlight that data analytics companies, such as Databricks or Integrate.io, misrepresent what they do as knowledge creation. Databricks promises to “[d]erive new insights.”⁵⁵ Integrate.io claims to be a “Single Source of Truth.”⁵⁶ In doing so, these companies present themselves as primarily in the business of knowledge production. However, as Julie Cohen points out, “[t]he data refinery is only secondarily an apparatus for producing knowledge; it is principally an apparatus for producing wealth. It facilitates new and unprecedented surplus extraction strategies within which data flows extracted from people . . . are commodity inputs, valuable only insofar as their choices and behaviors can be monetized.”⁵⁷ Businesses purchase data access because this allows them to plug consumers into their marketing goal for specific consumer segments.⁵⁸ Platforms and data analytic firms are not engaging in a benevolent form of knowledge production for knowledge’s sake. Instead, they are engaging in the profit-driven datafication that defines informational capitalism.

50. *Id.*

51. *See id.* at 69–70.

52. *See id.* at 70.

53. *Id.*

54. *See id.*

55. DATABRICKS, <https://databricks.com/> [<https://perma.cc/6WN7-KV2R>].

56. INTEGRATE, <https://www.integrate.io/> [<https://perma.cc/8USS-JPY3>].

57. *See* COHEN, *supra* note 3, at 71.

58. *See id.*

B. The Data Political Economy Also Produces Harms Based on Horizontal Relations

Not only is data's value predicated on the horizontal predictive relationship between data subjects, but current data practices enact horizontal harms on vulnerable populations. Again, a horizontal relationship is one between two data subjects, rather than a vertical relationship between user and platform. Viljoen critically argued that while everyone may be equally subjected to abusive vertical relations, like minimal agency in assenting to collection practices, horizontal relations between data subjects lead to unjust harms based on preexisting social inequality.⁵⁹ So while we all may have unconsented-to caches of information being collected on us, there is more harm when data becomes a conduit for preexisting societal discrimination to flow. A paradigmatic example of this horizontal harm is the Immigration and Customs Enforcement (ICE) and Customs and Border Protections (CBP) purchase of location data from data analytics companies.⁶⁰ These agencies purchased access to a commercial database that mapped the movements of millions of Americans.⁶¹ The location information located in the database had originally been collected by cellphone apps for games, weather, and e-commerce.⁶² The data was then used to locate and arrest undocumented immigrants based on the data points that placed them at remote places along the Mexican border.⁶³ Essentially, these government agencies used a large data cache to isolate features amongst the aggregated data that suggested undocumented status.⁶⁴ Because the authorities were able to purchase the data cache from marketing companies, the agencies were able to avoid the necessary warrant application they would have needed to gather the location information from cellphone companies themselves.⁶⁵

Another example is that of a top Catholic Church official, Jeffrey Burrell. He resigned after his cellphone data was used to show that he routinely used the gay dating app, Grindr, which tracked him to gay bars. While it is unclear who utilized the information on Burrell, someone was able to purchase a cache of data and use it to identify him. Whoever did this then reported the information to *The Pillar*, a newsletter that reports on the Catholic Church. This forthcoming news story led to Burrell's resignation. *The Pillar's* story reported that someone had

59. See Viljoen, *supra* note 23, at 614–15.

60. See Byron Tau & 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/633A-Y3R7>].

61. *Id.*

62. *Id.*

63. *Id.*

64. See *id.*; Viljoen, *supra* note 23, at 614–15.

65. See Viljoen, *supra* note 23, at 631 n.151; Tau & Hackman, *supra* note 60.

used purchased data to “correlate” Burril’s location at gay bars through his Grindr usage.⁶⁶

There are many more examples of abuse of location data, from the IRS using data to track suspects,⁶⁷ to apps that help Muslims time their prayers selling data to companies that then sell to military contractors,⁶⁸ to police using unreliable predictive algorithms based on crime data sets that reinforce the oversurveillance of Black, Brown, and poor communities.⁶⁹ Viljoen critically made clear that while the patterns that data reveal about us are not inherently oppressive, the data patterns that become associated with discriminated groups “become constitutive of how members of [these] population[s] are socially defined and acted upon in oppressive ways.”⁷⁰ These harms are horizontal because data analytic predications based on data sets of large populations become ways to target discriminated-against groups. In an already unequal society, data analytics can become tools to bolster preexisting structural inequalities.

C. *The Current Data Regulatory Structure Is Ineffective Against Both Vertical and Horizontal Harms*

The existing privacy framework does nothing to stop horizontal harms by focusing only on *individualized* data. The current data regulation regime is predicated on attempting to stop vertical harms between the data subject and the data cultivator, neglecting to consider the value that platforms derive from *aggregated* data. Many privacy laws are based on the Federal Trade Commission’s (FTC) Fair Information Practice Principles (FIPPS), which views fair data practices as those that give people meaningful control of how their data is processed and used.⁷¹ However, this has resulted in the regime of boilerplate contracts mentioned earlier, now known as “notice and consent,”⁷² that give an illusion of consumers being on notice to corporate data extraction practices. However, as articulated above in the boilerplate contracts discussion, notice and

66. See Michelle Boorstein, Marisa Iati & Annys Shin, *Top U.S. Catholic Church Official Resigns After Cellphone Data Used to Track Him on Grindr and to Gay Bars*, WASH. POST (July 21, 2021), <https://www.washingtonpost.com/religion/2021/07/20/bishop-misconduct-resign-burrill/> [https://perma.cc/Q6DR-BM76].

67. See Byron Tau, *IRS Used Cellphone Location Data to Try to Find Suspects*, WALL ST. J. (June 19, 2020), <https://www.wsj.com/articles/irs-used-cellphone-location-data-to-try-to-find-suspects-11592587815> [https://perma.cc/B5WG-N9FK].

68. See Joseph Cox, *More Muslim Apps Worked with X-Mode, Which Sold Data to Military Contractors*, VICE (Jan. 28, 2021), <https://www.vice.com/en/article/epdkze/muslim-apps-location-data-military-xmode> [https://perma.cc/8EU8-N2EM].

69. See Aaron Sankin, Dhruv Mehrotra, Surya Mattu & Annie Gilbertson, *Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them*, MARKUP (Dec. 2, 2021), <https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them> [https://perma.cc/NZJ9-P9PJ].

70. Viljoen, *supra* note 23, at 615.

71. *Id.* at 593.

72. *See id.*

consent is not a successful regulatory framework as it currently does not allow actual consent to the terms of data extraction and use.⁷³ In fact, platforms have many incentives to make enrollment in their data collection practices “seamless and near-automatic” to dissuade users from fully understanding what is at stake.⁷⁴ Again, this brings up Twitter’s and Facebook’s practices of automatically collecting data from people who go to a website with a “like” button linked to the respective platform.⁷⁵ Given the issues with current notice and consent structures, it is hard to argue that any current acquiescence to data collection protocols is meaningful within FIPPS guidelines.⁷⁶ Yet, as this Section has already shown, there are serious data harms that develop due to the relationality of data subjects. Even if notice and consent effectively empowered data subjects to meaningfully negotiate with data cultivators, this regulatory structure would do nothing to stop the horizontal harms because it only locates the possibility of harm in the vertical relationship. By only focusing on the flow of data from person to platform, FIPPS cannot adequately address the population-level harms the allowed for ICE’s use of data to target undocumented immigrants. Therefore, the issue with notice and consent is not only that it allows largely unconsented-to data harvesting, but that it also pools data into a resource that data analytic firms can process to enact societal systems on inequity.

Notice and consent regulation’s inability to deal with horizontal harms is epitomized by the story of Life360, a popular app. Life360 is a family safety app that has thirty-three million users worldwide. It is marketed as an app that allows families to know where other family members are at all times. It is particularly popular with parents who wish to track their children’s locations. The app provides precise real-time locations of users, and if they are in a car, the speed at which they are driving. While Life360 bills itself as a safety app for families, it makes a large percentage of its profits from selling its users’ location data to data brokers who have been known to sell data to the U.S. Department of Defense.⁷⁷ In 2020, Life360 made nearly 20% of its revenue from selling data culled from its userbase.⁷⁸

Notice and consent has done little to protect people from informational capitalism’s horizontal harms. Life360 is, in fact, one of many apps that sells into

73. See Sadowski, *supra* note 8, at 8.

74. See COHEN, *supra* note 3, at 58–59.

75. See David Ingram, *Facebook Fuels Broad Privacy Debate by Tracking Non-Users*, REUTERS (Apr. 15, 2018), <https://www.reuters.com/article/us-facebook-privacy-tracking-idUSKBN1HM0DR> [<https://perma.cc/QB47-WJCC>]; Dr_Jeff, *supra* note 15.

76. See Viljoen, *supra* note 23, at 593.

77. See Jon Keegan & Alfred Ng, *The Popular Family Safety App Life360 Is Selling Precise Location Data on Its Tens of Millions of Users*, MARKUP (Dec. 6, 2021), <https://themarkup.org/privacy/2021/12/06/the-popular-family-safety-app-life360-is-selling-precise-location-data-on-its-tens-of-millions-of-user> [<https://perma.cc/4FJ3-XE5G>].

78. *Id.*

the \$12 billion market for buying and selling location data.⁷⁹ With apps that use location features, it is hard to know which ones just use users' data for the app's functionality and which ones sell users' locations into the marketplace.⁸⁰

Once location data gets sold into the marketplace, "it can be sold over and over again, from the data providers to an aggregator that resells data from multiple sources."⁸¹ Apps, like Life360, use "seamless and near-automatic" consent that technically adheres to notice and consent.⁸² However, while still within current regulations, the selling and reselling of consumer data empowers data analytics to be deployed on large population data sets that enable the kinds of horizontal harm like ICE purchasing location data to target undocumented immigrants.⁸³ The failure of notice and consent does little to dissuade abusive data cultivation practices. Once firms already have the data, the current regime does nothing to protect people against the horizontal harms that data can inflict on the marginalized.

D. Proposed Data Reforms Also Fail to Address Horizontal Harms

Widespread recognition of the issues with the current data governance structure has led to two kinds of reform proposals. These two reform proposals fall under two categories: propertarian reforms and dignitarian reforms.⁸⁴ Both proposals fail because, like FIPPS, they fail to appreciate the horizontal nature of data's political economy.⁸⁵

Propertarian reforms argue that the issue with the political economy of data is that there is a formal absence of people's property rights in their data.⁸⁶ These proposals consider the activities that produce data as labor, and they are concerned that data subjects are not properly remunerated for their role in the wealth created by data.⁸⁷ An example of a propertarian reform is a legal structure that gives users a property-like claim to their data in exchange for payment and/or a private right of action to sue over misuse.⁸⁸

Dignitarian reforms involve regulation based on conceptions of human dignity that are undermined by data cultivation. These proposals view data as an extension of the human and are concerned about the dehumanizing effects of ceaseless datafication and commodification of our personal lives as inputs into

79. *Id.*; Jon Keegan & Alfred Ng, *There's a Multibillion-Dollar Market for Your Phone's Location Data*, MARKUP (Sept. 30, 2021), <https://themarkup.org/privacy/2021/09/30/theres-a-multibillion-dollar-market-for-your-phones-location-data> [<https://perma.cc/9C5P-X6DS>].

80. *See id.*

81. Keegan & Ng, *supra* note 79.

82. *See* COHEN, *supra* note 3, at 58.

83. *See* Viljoen, *supra* note 23, at 614–15.

84. *See id.*

85. *See* Salomé Viljoen, *Data as Property?*, PHENOMENAL WORLD (Oct. 16, 2020), <https://www.phenomenalworld.org/analysis/data-as-property/> [<https://perma.cc/M3SS-PVJN>].

86. *See id.*

87. *See id.*

88. *See id.*

informational capitalism.⁸⁹ Dignitarian reforms are also concerned with the algorithmic sorting of populations in the process of datafication, where data subjects become legible patterns with preferences that are fed into algorithms. In turn, the algorithms act on the individual, creating a positive feedback loop where “[the] 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.”⁹⁰ Dignitarian proposals advocate for legal regimes that treat data as an extension of the self and allow users to claim a right to self-determination outside of algorithmic sorting.⁹¹ As opposed to property that could be disposed of, such dignitarian natural rights in data would be seen as universal inalienable rights.⁹²

As Viljoen made clear in her Article, both proposals suffer from downfalls. However, both proposals also highlight important concerns of current data practices. Propertarian concerns are valid because it is true that people produce the data that makes platforms exceedingly wealthy without any benefit. Dignitarian concerns are additionally important because it is true that the continual data cultivation, surveillance, and subjection to algorithmic sorting can affect self-expression.⁹³ While both proposals raise valid concerns, neither proposal would adequately deal with the other proposal’s identified issue. For example, a propertarian regime would potentially give people money for the data produced, but it would undermine a dignitarian proposal because people would be incentivized to hand over greater amounts of data for money. Thus, a propertarian regime would contribute to the greater commodification of data that dignitarian reforms are concerned with.⁹⁴

However, most importantly, neither dignitarian nor propertarian proposals adequately deal with the horizontal harms of data collection. For example, under a propertarian legal regime, Person A could be incentivized to hand over greater amounts of data because there is remuneration for data production. However, Person B, who has a similar demographic and geographic profile to Person A, might make the conscious decision to not use any apps or produce as little data as possible. Person A’s data production would harm Person B because their similarity would in turn make Person B more recognizable to corporations despite an absence of Person B’s actual data. Therefore, Person B is not harmed through the vertical relation of a corporation harvesting their data, but they are instead harmed horizontally by Person A’s willingness to hand over data that makes Person B recognizable. A similar example could be made in a dignitarian regime. Let’s say again that Person B decides to invoke their inalienable

89. See Viljoen, *supra* note 23, at 623–24.

90. *Id.* at 624.

91. *See id.* at 625.

92. *See id.* at 626.

93. *See id.*

94. *See id.* at 622.

dignitarian right to be free from surveillance. However, Person A, who again has a very similar demographic and geographic profile, decides to opt in. Person B is again harmed horizontally by Person A's decision. These examples highlight precisely what Viljoen has argued: "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."⁹⁵ Both propertarian and dignitarian reforms suffer similar issues as FIPPS because they only target the vertical relationship structure of rights and obligations between the data subject and data cultivator. Therefore, both proposals fail to adequately address the horizontal relational structure that both harms people and drives data's value as a commodity.

Today, to be a part of society is to be digitized into data inputs for informational capitalism. This Section's argument reveals that the absolute power lies in the side of platforms that defines the political economy of data, while people who are requisite for data production have no meaningful ability to control their own data. Despite the fact that we produce tremendous value, we suffer scores of harms. These value and harms cannot be understood as individualized. Instead, as this Section argues, both the value and harms of data need to be understood in the horizontal relationships that produce both. This is not to say that the vertical harms considered by propertarian and dignitarian concerns are not important. However, these respective reform proposals are inadequate at dealing with the current political economy of data that is driven by the value created from horizontal relations. Instead of regulation that only deals with vertical harms, we need a data governance system that views "[d]atafication (or, more precisely, data production) [as] wrongful if and when it materializes unjust social relations along either the vertical or horizontal axis."⁹⁶

II.

THE NEED FOR INFORMATIONAL DEMOCRACY

In 1915, the United States Commission on Industrial Relations released a report that stated, "Political freedom can exist only where there is industrial freedom; political democracy only where there is industrial democracy."⁹⁷ The demand for "industrial democracy" galvanized the labor movement of the early 20th century.⁹⁸ Industrial democracy was viewed as a solution to the lack of control that workers experienced at work. It was a rallying cry "for 'the absolute and inalienable rights of workers' to exercise 'a compelling voice' in determining their working conditions."⁹⁹ The right of workers to organize in

95. *Id.* at 628.

96. *Id.* at 631.

97. MCCARTIN, *supra* note 17, at 12.

98. *See id.* at 12–13.

99. *Id.* at 27.

unions to demand a more just workplace was critical to the foundation of industrial democracy.¹⁰⁰

In today's world of informational capitalism, we need *informational democracy* to be at the center of any data regulatory regime. Our lives are increasingly dictated by algorithms we understand little about,¹⁰¹ and platforms tinker with us as experimental subjects to figure out how to nudge our actions in certain ways.¹⁰² People, constructed as data subject inputs into informational capitalism, are constrained and manipulated by systems designed for extraction and profit. As industrial democracy was about the need for workers to have greater say in how the workplace operated, this Note's invocation of informational democracy is about people's ability as data producers to meaningfully exert control over the data economy that their daily lives engender. Any regulatory framework that has informational democracy at its core must affirmatively empower citizens to meaningfully control data's vertical and horizontal relationships on their own terms.

Labor unions are an inspiration for how informational democracy can be implemented today. In the workplace, unions are critical to bringing material change on workers' terms. Because individual workers lack the necessary power to bring management to the table for meaningful negotiation, unions allow for workers to aggregate their influence together to collectively demand better terms.¹⁰³ Like workers in the workplace, data subjects' value is based on their aggregate contribution, and individual data subjects lack any meaningful ability to negotiate with platforms.¹⁰⁴ Based on the similar need of a large unit for bargaining, the key governance structure for industrial democracy, the union, would also work for informational democracy. The consolidation of people's data into large union data pools would, as unions give collective bargaining power in negotiating with management, allow unionized data subjects to take meaningful control back from the platform power that increasingly shapes the world.

This Section will first show what unions have traditionally accomplished in the labor context. Second, it will describe how data unions could be organized. Third, it will show how data unions, as an informational democracy governance tool, would deal with the data economy's vertical and horizontal harms.

100. See *id.* at 28.

101. See, e.g., PASQUALE, *supra* note 12.

102. See, e.g., 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) (describing an experiment that tested the effects of political mobilization messages on sixty-one million-people disseminated over social media).

103. See Kim Kelly, *What a Labor Union Is and How It Works*, TEEN VOGUE (Mar. 12, 2018), <https://www.teenvogue.com/story/what-a-labor-union-is-and-how-it-works> [https://perma.cc/9XTB-TLR3].

104. See Sadowksi, *supra* note 8, at 7–8.

A. Unions Increase Wages and Protect Worker Dignity

Unions are bargaining units with democratically elected leadership who advocate for workers to have increased wages and promote workers' ability to live dignified lives.¹⁰⁵ Unions help workers to receive a larger share of the value that their work produces. On average, unionized workers have salaries that are 28 percent higher than similarly situated nonunionized workers.¹⁰⁶ Furthermore, despite the racist legacy of many unions in the 20th century,¹⁰⁷ today, union membership for Black workers is a key factor in lowering the Black-White wage gap. This is because Black workers are more likely than White workers to be in unions, and Black workers get a larger boost to their wages from being in a union than White workers do. Black workers get paid 13.1 percent more than similarly situated nonunionized Black workers. These same statistics apply to unionized Latinx workers, who get paid 20.1 percent more than their nonunionized peers.¹⁰⁸ Furthermore, being a union member was critical for many workers' ability to weather the economic crisis precipitated by COVID.¹⁰⁹

Unions are not just successful in increasing wages, but they are also essential for protecting workers' ability to live a dignified life. Since the early 20th century, workers have used the phrase "bread and roses" as a metaphor for the working conditions they demanded.¹¹⁰ This slogan signified that workers did not want just the needed wage to buy "bread" for survival, but that they also wanted working conditions that would allow them to enjoy the "roses" of life.¹¹¹ Banding together in unions aided workers in finding "a voice in determining the conditions of their work, and their desire to claim their rights as citizens through their labor."¹¹² Unions have intangible benefits giving workers the "dignity and independence that comes when 'they have to treat you like a [person].'"¹¹³ The dignitarian concerns for unions can be seen today with wins that give workers meaningful paid vacation, consistent schedules that allow workers to schedule

105. *What Unions Do*, AFL-CIO, <https://aflcio.org/what-unions-do> [<https://perma.cc/9LKT-2JPD>].

106. Kelly, *supra* note 103.

107. *See, e.g.*, RICHARD ROTHSTEIN, *THE COLOR OF LAW: A FORGOTTEN HISTORY OF HOW OUR GOVERNMENT SEGREGATED AMERICA* 158, 160–61 (2017) (describing the practice of unions forcing companies to fire African Americans, unions refusing to accept African Americans, and unions depriving African Americans of benefits enjoyed by their white counterparts).

108. Celine McNicholas, Lynn Rhinehart, Margaret Poydock, Heidi Shierholz & Daniel Perez, *Why Unions Are Good for Workers – Especially in a Crisis Like COVID-19*, *ECON. POL'Y INST.* (Aug. 25, 2020), <https://www.epi.org/publication/why-unions-are-good-for-workers-especially-in-a-crisis-like-covid-19-12-policies-that-would-boost-worker-rights-safety-and-wages/> [<https://perma.cc/P84B-AE4Y>].

109. *See generally id.* (describing how protections won by unions serve to protect workers during crises, like COVID).

110. *See* Robert J.S. Ross, *Bread and Roses: Women Workers and the Struggle for Dignity and Respect*, 16 *WORKINGUSA* 59, 59–60 (2013).

111. *See id.* at 59.

112. MCCARTIN, *supra* note 18, at 95.

113. ANDREW LEVISON, *THE WORKING CLASS MAJORITY* 180 (1974).

their lives, and the increased likelihood of obtaining health insurance and paid family leave.¹¹⁴

While the labor context obviously differs from data production, the discussion above highlights that labor unions effectively achieve goals that their members could not reach individually. The organization of worker unions was a response to the lack of bargaining power that laborers possessed vis-à-vis their employers. Individually, the worker had no ability to meaningfully bargain with their employer regarding the terms of their employment. Yet, their collectivism forced employers to the table. As this Note states, data subjects similarly lack any ability to bargain with platforms to the terms of their data usage. Because data in the aggregate creates its value, the collective withholding of data access would meaningfully tilt the current power imbalance between data subjects and data cultivators.

B. For the Data Unions Make Us Strong

Data unions would nurture the growth of informational democracy. The premise of the data union is inserting a point of friction in the initial extractive vertical relationship between data subject and data cultivator. Data unions would be a new third party directly between data subjects and the platforms as data cultivators. Instead of data flowing uninhibited to the cultivator, the data would be shunted off into the data union. As previously noted, current data collection practices engender no agency on the part of the person whose data is collected.¹¹⁵ It is at the point of initial contact between the data subject and cultivator at which the person loses any control over their data. However, in a data union regulatory structure, platforms and other entities would no longer be entitled to user data because the data would first be deposited in a union the user is a member of. Therefore, if any entity desired access to data, it would be forced to negotiate the terms of access with the data union. Such collective bargaining with platforms has the potential to radically clear the opaqueness with which many platforms currently operate and use data.

Essential to the organization of data unions would be the cornerstone that data unions “conceive of citizen data as a public resource (or infrastructure) to be managed via public governance and in furtherance of public goals.”¹¹⁶ This paper conceives of data unions as democratic institutions with regularly elected leadership that implement the general policy goals of the whole union. However, beyond these initial foundational points, there is the question of how to organize data into union structures. Labor unions are organized by workplace and trade. Workers know which union to go to based on their employer and the nature of their work. What about for data collectives?

114. See *What Unions Do*, *supra* note 105; Kelly, *supra* note 103.

115. See COHEN, *supra* note 3, at 44.

116. Viljoen, *supra* note 23, at 646.

There are a couple of possibilities for how data unions could be organized. One way to organize data unions is by geographic locality. The union could exist at any geographic size, from state, county, city, or even neighborhood. Union membership dictated by geography is well-suited to deal with the harms caused by location data mining. Location data is highly valuable for many reasons, as seen earlier with the examples of government purchases to target certain groups.¹¹⁷ It is also incredibly valuable because hedge funds use it to understand foot traffic of certain businesses to guide their investment decisions.¹¹⁸ Geographically-bounded data unions based on location would best deal with such harms because, by binding together a location's data, whole areas of location data would be uniformly affected.

To understand why geographic locality unions would best serve location data harms, we must imagine an example of unions not based on geography. Non-geographic unions could potentially have different views on location data, which would mean data cultivators have varying access to data. In this hypothetical, there are two neighbors, A and B. A is in a union with a lax data policy surrounding location. A's union essentially hands over any location data that a data cultivator seeks out. B is in a union that tightly controls location data. As next-door neighbors, A and B live very similar lives. Despite B's best efforts to be in a union that closely guards location data, firms and data analytics would still have access to A's movements. As previously stated, data's value is not predicated on the individual's data but on what it predicts about similarly situated groups. Despite a firm's lack of access to B's data, a firm could still build a profile on B with A's movement data because A and B are next-door neighbors. Thus, B's best efforts to be shielded from such practices are undermined and B would still be horizontally harmed by A's data. As this hypothetical makes clear, two neighbors, even with different data protections in place, can undermine the very purpose of data unions as regulatory structures. This is because despite B being in a union with strict rules on what types of location data can be passed onto platforms, A's information still has the potential to harm B's privacy interests.

Data unions bounded by geographic limits would allow communities to decide how data policies would best serve themselves. This form of data union could be an important way that marginalized communities employ self-determinative practices. Today, the United States is still highly segregated by race and class.¹¹⁹ As stated before, data practices and the predictions they enable are not in and of themselves discriminatory. The issue is that in an unequal society, data then becomes a tool that magnifies targeted oppression of already-

117. See, e.g., Keegan & Ng, *supra* note 77; Viljoen, *supra* note 23, at 614–15.

118. See Keegan & Ng, *supra* note 77 (“Companies like real estate firms, hedge funds and retail businesses might then turn and use the data for their own advertising, analytics, investment strategy, or marketing purposes.”).

119. See Sankin et al., *supra* note 69.

marginalized groups.¹²⁰ As neighborhoods in the United States are still often segregated by class and race, underlying discrimination allows for data practices to manifest as violent physical intrusions into such communities.

Consider, for example, police departments' usage of predictive policing software. Despite known inaccuracies in the predictions, Predpol, an algorithmic software, relentlessly targeted Black and Latinx neighborhoods and communities that qualified for federal free and reduced lunch programs.¹²¹ For many White and more middle- to upper-class neighborhoods, Predpol went years without predicting a single crime.¹²² Data-driven predictive policing is almost impossible to achieve because Black and Brown community members are more likely to report crimes than White and upper-class community members. It is not that crimes do not happen in White and upper- and middle-class neighborhoods, but that less reports of such crimes are completed.¹²³ Despite PredPol's acknowledgment of studies that showed its algorithm reinforced inequality, PredPol still went ahead and marketed its program to police forces across the country.¹²⁴ Geographic unions would be an advantageous bargaining unit because who knows better than the communities themselves as to what issues and discrimination they are facing? In a world of Predpol software, communities would be empowered to withhold any data on crimes (especially if the data union policy's covered local policing institutions) or data that would allow algorithms to classify their neighborhoods as poor, working class, Black, Latinx, immigrant, or similar neighborhoods. Geographic unions would be informational democracy tools that empower discriminated-against communities to have a material say in how their data affects their community.

Rather than discrete and separate geographical unions, it could make sense to have one national data union that combines all U.S. residents into one pool that is broken down into various geographical subunits. This could look like one national pool with data pools broken down into geographic regions like Northeastern states, West Coast states, and Midwestern states. The next level down would be the data union for the state and so on. Each geographical unit would be empowered by its own bargaining unit that sets policy. Such a structure would operate in concentric circles of policy: policies that are voted on and win at the national level apply nationally then data governance structures that win at the regional level then apply regionally. This may be more desirable than separate geographic unions spanning the country, like for every state or city, as this could lead to a splintered and ungovernable web of data policies across the country. A national-subnational structure instead could create more uniformity. Under this national-subnational union structure, smaller union units could still

120. See Viljoen, *supra* note 23, at 615.

121. See Sankin et al., *supra* note 69.

122. See *id.*

123. See *id.*

124. See *id.*

have differing policies, while large areas of the country would be governed by uniform policies.

The scope and increasing population of each level would also help ensure adequate bargaining power at various levels of size. If, for example, unions were just tethered to a state or city, then depending on population, different localities would have different power to bargain with platforms over data usage. A New York City data union, due to its population and size of its aggregated data, would have much more bargaining potential for favorable policy than a small town would have. By structuring unions in a national system of decreasing size, a small-town data union in New York state would be tied into the next higher level of organization, the New York state's data union. This would allow for great population density, and therefore greater bargaining potential. A smaller New York town, at a higher level of geographic organization, could then be tied into policy that the population from New York City can help them win because they would all be folded into the New York state data union.

Instead of organizing data unions based on geography, the organization of a data union could be predicated on policy. In this scenario, there could be several unions available to U.S. residents nationally and, based on each union's policy platforms regarding data, people could decide which union they want to be a part of. This would have the benefit of ensuring that every person's data is largely handled in a way that best accords with their views on data practices. For instance, if we had geographical data bargaining units and if someone lived in an area where the population overall did not agree with that person's views on data practices, then that person would still be subjected to data cultivation that undermined their beliefs. In data unions predicated on policy distinction, everyone would be able to have their data handled largely in line with how they wish. Yet, organizing data union structure around policy could be undesirable due to the lack of cohesive locality information. This point goes back to the argument raised in the example above of next-door neighbors A and B, where A's legibility of location information impacts B despite B's union's tight control of location data. Therefore, this kind of horizontal harm would not be fully protected against without unions based on geographic units.

C. Data Unions Could Protect Against Vertical and Horizontal Harms

Unions as data governance tools would protect people from both vertical and horizontal harms of data cultivation practices. Importantly, data unions would enable people to deal with the concerns that are highlighted by propertarian and dignitarian reform proposals.

Dignitarian concerns claim that data is an extension of human self-expression and are concerned with the ways that data cultivation practices undermine the self's autonomy. One such dignitarian objection to current data regimes, highlighted earlier, is to the feedback loop that develops when inputting predictive behavior as data doubles into algorithms creates a cycle where such

predictions become reinforced as actions are brought upon the people tied to the prediction.¹²⁵ This concern is explored in depth in Professor Frank Pasquale's book, *The Black Box Society: The Secret Algorithms That Control Money and Information*, which argues that society is more and more controlled by algorithms that are shrouded in secrecy.¹²⁶ This process of secretive algorithmic sorting touches processes from welfare benefit determinations to credit scoring to foster care determinations.¹²⁷ In recent years, legal aid lawyers have been flagging a growing number of public services that are determined through algorithms.¹²⁸ However, government agencies that utilize such algorithmic sorting do so without transparency or accountability. For example, a direct services lawyer attempted to challenge a determination for a client who had been cut off from Medicaid. The nurse representing the government at trial was unable to explain why the client had been taken off Medicaid because the determination had been made by an algorithm that the nurse did not understand, nor did she have access to its determination process.¹²⁹

Furthermore, as more state and city governments across the country switch to using algorithms for determining access to public benefits, their respective legislatures have little to no understanding as to how these algorithms work and which services are being wed to them.¹³⁰ To make matters worse, any attempts by legislatures to pass bills aimed at better understanding government use of algorithms have been met by antagonistic tech lobbying campaigns to stall such bills.¹³¹ Thus, government's basic services are contracted out to AI corporations that opaquely exert control over vulnerable populations' access to essential services without any effective oversight whatsoever. While the poor are uniquely vulnerable, algorithms affect us all, from how Google lists search results to how credit scores are calculated.¹³² Dignitarian data reforms are concerned by the unaccountability over these algorithms that shape our everyday lives.

Data unions would be an effective tool to remedy the ways our lives are unaccountably shaped by algorithms. In data unions' negotiations with government agencies or platforms that want unionized data to be placed it into algorithms, data unions could demand that algorithms' criteria and innerworkings be revealed. Such a bargain would allow data unions to uncover and make public the ways that private algorithms are ordering life and help

125. See Viljoen, *supra* note 23, at 624.

126. See PASQUALE, *supra* note 12.

127. See Hao, *supra* note 20.

128. See *id.*

129. See *id.*

130. See Todd Feathers, *Why It's So Hard to Regulate Algorithms*, MARKUP (Jan. 4, 2022), <https://themarkup.org/news/2022/01/04/why-its-so-hard-to-regulate-algorithms> [<https://perma.cc/8VL8-P5CV>].

131. See *id.*

132. See PASQUALE, *supra* note 12, at 23, 59; Rose Eveleth, *Credit Scores Could Soon Get Even Creepier and More Biased*, VICE (June 13, 2019), <https://www.vice.com/en/article/zmpgp9/credit-scores-could-soon-get-even-creepier-and-more-biased> [<https://perma.cc/8CSC-XBWP>].

illuminate how the world is being shaped. With such information out in the open, the public would be better able to contest such sorting. In fact, once algorithmic processes become known, data unions could in turn refuse to deal with any platform or government agency that uses algorithms shown to impose ineffective or discriminatory classifications on people. There are also growing movements to suggest that platforms should be banned from using surveillance advertising to target specific demographics of people.¹³³ Unions would be empowered to decide that they will no longer sell data to marketing or data analytic firms, or that they would only provide data to corporations that did not engage in surveillance marketing. There are endless possibilities for how data unions could determine what kinds of algorithms can access their data and on what terms.

Propertarian reforms are those that wish to provide data subjects formal, legal property rights over their data. In this way, people would be able to monetize or have effective control over their data once it is produced through legal ownership.¹³⁴ The treatment of aggregated data as the common property of the data union would also address propertarian concerns. By allowing property in data for the union, the union would be able to support itself economically with the underlying value of the data. As the union negotiates and bargains with platforms and other entities for access to the union's data, the contract could include a term specifying that a fee must be paid to the union for access to certain pieces of data. By utilizing the data as value-raising property, the unions' cost would be covered, empowering the bargaining unit to do the work of effectively protecting members' interests. This is important because it is unlikely that individualized payments to data subjects are possible due to complexity of a micro-payment system based on data production.¹³⁵ However, the union itself would have access to the aggregate value of its members data, and it could gain value from its negotiation. In this way, while people would be unable to receive monetary payment, they would still be able to benefit materially because the union, as a self-sustaining monetary entity, would have the ability to achieve informational democracy. This would also avoid the previously mentioned concern regarding propertarian reforms that monetizing data would merely encourage people to hand over more data, increase extraction from daily life, and undermine dignitarian concerns.

Most importantly, beyond vertical propertarian and dignitarian concerns, data unions would also effectively deal with horizontal harms. The value gained from predicative analytics on aggregate data drives the political economy of

133. See Press Release, Accountable Tech, Dozens of Key Stakeholders Formally Urge FTC to Ban Surveillance Advertising (Jan. 27, 2022), <https://accountabletech.org/media/dozens-of-key-stakeholders-formally-urge-ftc-to-ban-surveillance-advertising/> [<https://perma.cc/TNV8-3492>]; *FTC Explores Rules Cracking Down on Commercial Surveillance and Lax Data Security Practices*, FTC (Aug. 11, 2022), <https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-explores-rules-cracking-down-commercial-surveillance-lax-data-security-practices> [<https://perma.cc/M9GW-A7RU>].

134. See Viljoen, *supra* note 85.

135. See *id.*

data.¹³⁶ Therefore, to be an effective tool for harnessing meaningful informational democracy, data governance reforms must be in tune with how data creates value and harms. Data unions would be responsive to the horizontal harms. With data unions giving aggregated groups of persons control over their aggregate data, corporations could be disempowered from making the demographic-specific predictions necessary for rendering people data-legible. Most importantly, by grouping data subjects together, unionized members would have the same data protocols applied to their data. Let's return to the example used earlier. Person B does not want to produce any data, but is demographically similar to Person A in location, age, income range, race, etc. As much as Person B may try to protect themselves from corporate legibility through data practices, Person A's data could still be used to make predictions about Person B. However, if Person A and Person B were instead bound by the same data regulatory rules because of a geographic union, Person A's data decision would no longer have as strong an impact on Person B because they would both be under the same regulatory structure. Thus, it would no longer just be up to Person A as to how their data gets used. Instead, as equal union members in a data union, Person B also has a say as to how Person A's data gets used and impacts Person B.

Take the real-world example of ICE purchasing access to data pools to target undocumented immigrants—by passing data to a union first, the union could decide that location data will never be passed onto a third party. Even if a union did decide to pass along some location data, the union would be empowered in negotiating to demand only certain uses for data. Such negotiations could include refusing to provide data of any kind to criminal enforcement offices, immigration agencies, or defense contractors. As for the initial and control repository of data, the union would be in the position to decide what outside access to data consists of. If data unions are democratically governed, people will be empowered to decide what such terms will look like. In this way, the critical need for the greater control of informational capitalism, reined in by informational democracy, can be achieved.

CONCLUSION

The economy of data production can function as it does only because platforms have the sole power to define the terms of their access to our data, while we, the data producers, have no say whatsoever. This arrangement has awarded tech corporations with immense wealth and tremendous power that allows them to unaccountably shape our daily lives. Yet despite the status quo, things need not remain the same. As author Ursula K. Le Guin once said, “We live in capitalism, its power seems inescapable – but then, so did the divine right of kings. Any human power can be resisted and changed by human beings.”¹³⁷

136. See COHEN, *supra* note 3, at 70.

137. Ursula K. Le Guin, *Ursula K Le Guin's Speech at National Book Awards: 'Books Aren't Just Commodities,'* GUARDIAN (Nov. 20, 2014),

With a nod to Le Guin's description of the malleability of human power, this Note argues that even though the tech titans currently exercise invasive control, our power, the combined efforts of citizen data producers, can bring these corporations to heel. As the laborers of the late 19th and early 20th century collectively stood against the industrial barons of their time in calling for industrial democracy, so too must data producers join together today to demand informational democracy: the ability to meaningfully exert control over the data economy that our lives engender. As this Note has shown, because of the horizontal nature of both the data economy's value and harms, data unions would be a unique tool to handle regulation over the datafied economy.

While this paper suggests a new way of envisioning democratic and collective data regulation, it is only the starting point of imagining a future where citizens meaningfully engage with the data economy. It is the goal of this paper to argue that data unions would be an effective tool, yet this argument in turn raises other questions. For example, how would data unions be implemented? Would they need to be enacted by congressional statute that created the quasi-governmental structure that allowed for the geographical union structure tethered to locality? Or could individual citizens themselves organize data unions as apps or platforms embedded in their devices that exerted control over user data before other data cultivators could access them? Another strain of questions asks whether everyday citizens are literate enough in understanding data practices to vote on policies if data unions existed today. Current studies suggest that individuals might currently not have the requisite knowledge of data to be considered literate. Thus, questions of how to best educate populations about data and whether data unions could simultaneously become a tool to teach broadly about harmful data practices remain.¹³⁸ While beyond the scope of this initial paper, these are all important questions to be answered someday.

If we address the need for everyday people to meaningfully thwart data harms, the status quo will not continue. Our likes, our fears, our desires, and our lives create the data. We create the value in this informationalized economy. We imbue the tech giants with their tremendous power. We need informational democracy that democratically balances this inequity of power. Data unions are one such way.

<https://www.theguardian.com/books/2014/nov/20/ursula-k-le-guin-national-book-awards-speech>
[<https://perma.cc/A2CT-3QSG>].

138. See Simeon Yates & Elinor Carmi, *Don't Know How Your Data Is Used, or How to Protect It? You're Not Alone – But You Can Improve Your Data Literacy*, CONVERSATION (Oct. 11, 2021), <https://theconversation.com/dont-know-how-your-data-is-used-or-how-to-protect-it-youre-not-alone-but-you-can-improve-your-data-literacy-169431> [<https://perma.cc/AUL5-MEMB>].