

AN ECOLOGICAL APPROACH TO DATA GOVERNANCE

JASMINE MCNEALY*

I. INTRODUCTION

The products of algorithmic and decision systems unleashed in the wild—put to use by governments, corporations, and civil society organizations—significantly impact how life happens and society functions. Recently, a Facebook whistleblower detailed how the inability to recognize these significant impacts, and the failure of federal lawmakers to protect people against them, has led to an exponential concentration of power.¹ Machine learning systems have produced biased outcomes in consequential areas like school admissions,² government services,³ financial services,⁴ and healthcare.⁵ Algorithmic social media failures influence public opinion about governance, public health, and body image.⁶ Research demonstrates that distrust in automated systems is related to individual perceptions about whether a system can do what it promises to do.⁷ Therefore, when algorithms produce outcomes

* Associate Professor, University of Florida. Much appreciation goes to the helpful reviewers and participants at the Privacy Law Scholars Conference.

1. See John D. McKinnon & Ryan Tracy, *Facebook Whistleblower's Testimony Builds Momentum for Tougher Tech Laws*, WALL ST. J., <https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-set-to-appear-before-senate-panel-11633426201>, (Oct. 5, 2021, 5:21 PM).

2. See D. J. Pangburn, *Schools Are Using Software to Help Pick Who Gets In. What Could Go Wrong?*, FAST CO. (May 17, 2019), <https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong>; Oscar Schwartz, *Untold History of AI: Algorithmic Bias Was Born in the 1980s*, IEEE SPECTRUM (Apr. 15, 2019), <https://spectrum.ieee.org/tech-talk/tech-history/dawn-of-electronics/untold-history-of-ai-the-birth-of-machine-bias>.

3. See Faith Gordon, Book Review, 1 L. TECH. HUMANS 162, 162 (2019) (reviewing VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (2018)), Rebecca Heilweil, *Why Algorithms Can Be Racist and Sexist*, VOX (Feb. 18, 2020, 12:20 PM), <https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency>.

4. See Jennifer Miller, *Is an Algorithm Less Racist Than a Loan Officer?*, N.Y. TIMES (Sep. 18, 2020), <https://www.nytimes.com/2020/09/18/business/digital-mortgages.html>; Michelle Seng Ah Lee & Luciano Floridi, *Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-offs*, 31 MINDS & MACHS. 165 (2021).

5. See Heidi Ledford, *Millions Affected by Racial Bias in Health-Care Algorithm*, 574 NATURE 608 (2019); Eliza Strickland, *Health Care Algorithms Show Racial Bias*, IEEE SPECTRUM, Jan. 2020, at 6 (2020).

6. See McKinnon & Tracy, *supra* note 1.

7. See Mary T. Dzindolet et al., *The Role of Trust in Automation Reliance*, 58 INT'L J. HUM.-COMPUT. STUD. 697 (2003); Jiun-Yin Jian et al., *Foundations for an Empirically Determined Scale of Trust in Automated Systems*, 4 INT'L J. COGNITIVE ERGONOMICS 53 (2000); John O'Donovan & Barry Smyth, *Trust in Recommender Systems*, in IUI '05: PROCS. OF THE 10TH INT'L CONF. ON INTELLIGENT USER INTERFACES 167 (2005), <http://portal.acm.org/citation.cfm?doid=1040830.1040870>.

with tremendous, disparate, negative impacts, it results in calls for some kind of course correction.

Current such calls reflect the concern of many that there is a limited understanding of how a particular system works and comes to its decision. Demands for transparency and explainability in algorithmic systems are increasing, even spilling into the courts. In 2017, for example, the United States Supreme Court denied an appeal from a state supreme court hoping to force algorithmic transparency under constitutional grounds.⁸ In *Loomis v. Wisconsin*, a criminal defendant argued that the use of the COMPAS risk assessment tool at sentencing violated his right to due process “either because the proprietary nature of COMPAS prevents defendants from challenging the COMPAS assessment’s scientific validity, or because COMPAS assessments take gender into account.”⁹ The Wisconsin Supreme Court ultimately concluded that using the risk assessment tool for sentencing does not violate due process “if used properly, observing the limitations and cautions.”¹⁰ This decision has been panned as a threat to constitutional due process.¹¹ It also illustrates the need for accountability for algorithms used in governance.¹²

Loomis reflects a concern about the black box nature of algorithmic tools and systems. Black boxes are those systems “colonized by the logic of secrecy,”¹³ having practices invisible to humans, and yet having significant impacts on our lives and the environments—legal, financial, social—that we inhabit. An “opening” of the black box requires more than just knowing the algorithmic processes, but understanding them as well, “doing neuroscience” as some have called it.¹⁴ Yet, legislators are considering how to require explainability for algorithmic systems; the right to explanation is included, for instance, in the EU’s General Data Protection Regulation (GDPR).¹⁵ Some argue, however, that explainability and transparency may be no match for algorithmic complexity.¹⁶

8. *Loomis v. Wisconsin*, 137 S. Ct 2290 (2017) (*cert. denied*).

9. *State v. Loomis*, 881 N.W. 2d 749, 753 (Wis. 2016).

10. *Id.* at 753.

11. See Katherine Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, 18 N.C. J. L. & TECH. 75 (2016); John Lightbourne, *Damned Lies & Criminal Sentencing Using Evidence-Based Tools*, 15 DUKE L. & TECH. REV. 327 (2017).

12. See Han-Wei Liu et al., *Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability*, 27 INT’L J. L. & INFO. TECH. 122 (2019).

13. FRANK PASQUALE, *THE BLACK BOX SOCIETY* 2 (2015).

14. Davide Castelvecchi, *The Black Box of AI*, 538 NATURE 20 (2016).

15. But see Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. & TECH. 841 (2018); Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT’L DATA PRIV. L. 76 (2017) (arguing that the GDPR does not actually create an explainability requirement).

16. See Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889, 891 (2017).

One way to understand complex ideas is through analogy. If, for example, algorithms are analogized to recipes,¹⁷ they assist in illustrating that these systems are programmed to use “raw” materials to produce other materials—decisions and/or predictions. Algorithms, as representation systems, provide “different ways of organizing, clustering, arranging and classifying concepts, and of establishing complex relations between them” and allowing for the production of correlations, or the formulation of (perceived) relationships.¹⁸ And although this analogy may not completely identify or allow for understanding the exact ways AI systems produce their decisions, it recognizes that these systems are based on data.¹⁹ Because of this, a critical consideration of the data used to feed algorithms and how it is governed may be a way to circumvent the black box nature of these systems, and to obtain a better understanding of how they operate.

This essay offers a critical investigation of data and how it should be defined and governed to produce more transparency and mitigate possible harms to individuals and communities because of its use in AI systems. In essence, this essay argues that data should be viewed as a networked representation or observation. This definition recognizes that data is not singular, but always comes attached with labels, contexts, and biases fastened from its inception, if not collection, and that attachments increase depending on its place in the ecosystem. This view also requires a different strategy for governance—one that acknowledges data’s nature and networked existence, and moves beyond the individualistic, consent-based current models. Such an approach allows for the creation of better frameworks for collection, use, storage, access, and security of data. At the same time, this writing lays out a research agenda for further exploration of frameworks for harm reduction.

II. (RE)DEFINING DATA

Data is governed based on how it is imagined and is defined. Property language and the rhetoric of ownership in relation to personal data, create a description of data divorced from individuals and ignore the potential for harm. Phrases like the colloquial “my data” or “their data,” depict data as though they were a singular, unattached object. The ongoing controversy of the increasing use of genetic ancestry databases by law enforcement may help to illustrate the potential for harm in this kind of understanding of data. Over the past few years, the popularity of DNA matching and ancestry information has grown, spurred, no doubt, by the speed of DNA sequencing and the relative inexpensiveness of

17. See Kevin D. Ashley & Edwina L. Rissland, *Law, Learning and Representation*, 150 A.I. 17 (2003); M. Galieh Gunagama, *Generative Algorithms in Alternative Design Exploration*, SHS WEB CONF., VOL. 41 2018, at 2.

18. Stuart Hall, REPRESENTATION: CULTURAL REPRESENTATION AND SIGNIFYING PRACTICES 15, 17 (1997).

19. See CATHY O’NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); Gordon, *supra* note 3.

home testing.²⁰ This has meant the expansion of genetic testing outside of healthcare and paternity contexts, and into homes. Further, the popularity of ancestry-focused media content like, “Who Do You Think You Are?” and “Finding Your Roots” further drive interests in genetic testing and ancestry.²¹

The commercial genetic testing industry in the United States has grown immensely, dominated by players like Ancestry.com, 23andMe, and GEDMatch. Individuals can purchase at-home testing kits that require them to send in a quantity of saliva or a cheek swab by mail. These are then tested for connections to samples already in the organization’s database, returning possible results about heritage, ethnic identity, and long-lost relatives. There is a genre of YouTube and other social media posts in which users reveal their genetic ancestry results, expressing ranges of emotions from surprise to anger to confusion.²²

While the popularity of direct to consumer DNA ancestry testing has been interesting for individuals, the possibilities of this database has not gone unnoticed by law enforcement officials.²³ In fact, over the past few years, the number of law enforcement requests for access to the data stored by these organizations has increased, no doubt spurred by the use by law enforcement in California to catch the famed Golden State Killer (GSK).²⁴ In that case, police requested access to the site GEDMatch to test against a sample of DNA left by the alleged killer at a crime scene. With access, law enforcement found someone who matched with a percentage of the police sample, indicating a “cousin” relationship.

The relative success of the GSK case has had reverberations both for law enforcement and the commercial DNA testing industry. First, law enforcement

20. See JENNIFER KING, “BECOMING PART OF SOMETHING BIGGER”: DIRECT TO CONSUMER GENETIC TESTING, PRIVACY, AND PERSONAL DISCLOSURE, *PROC. ACM HUM.-COMPUTING INTERACTION*, Nov. 2019, at 158:1; Antonio Regalado, *2017 Was the Year Consumer DNA Testing Blew Up*, MIT TECH. REV. (Feb. 12, 2018), <https://www.technologyreview.com/2018/02/12/145676/2017-was-the-year-consumer-dna-testing-blew-up/>.

21. See Ashley Barnwell, *The Genealogy Craze: Authoring an Authentic Identity through Family History Research*, 10 *LIFE WRITING* 261, 262 (2013); Wendy D. Roth & Biom Ivemark, *Genetic Options: The Impact of Genetic Ancestry Testing on Consumers’ Racial and Ethnic Identities*, 124 *AM. J. SOCIO.* 150 (2018).

22. See Anna Harris et al., *Autobiologies on YouTube: Narratives of Direct-to-Consumer Genetic Testing*, 33 *NEW GENETICS & SOC’Y* 60 (2014).

23. See Claire Abrahamson, *Guilt by Genetic Association: The Fourth Amendment and the Search of Private Genetic Databases by Law Enforcement*, 87 *FORDHAM L. REV.* 50 (2019); Christi J. Guerrini et al., *Should Police Have Access to Genetic Genealogy Databases? Capturing the Golden State Killer and Other Criminals Using a Controversial New Forensic Technique*, *PLOS. BIOL.*, Oct. 2018, at 1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6168121/>; Rachele M. Hendricks-Sturup et al., *Direct-to-Consumer Genetic Testing and Potential Loopholes in Protecting Consumer Privacy and Nondiscrimination*, 321 *J. Am. Med. Assoc.* 1869 (2019); Joseph Zabel, *The Killer Inside Us: Law, Ethics, and the Forensic Use of Family Genetics*, 24 *U.C. BERKELEY J. CRIM. L.* 47 (2019).

24. See Guerrini et al., *supra* note 23; Zabel, *supra* note 23.

are now seeking further access to these databases.²⁵ This “crisis” has caused lawmakers in Utah to propose legislation that would deny law enforcement access to these databases.²⁶ Though such a proposal coming out of Utah may seem strange on its face, it is important to note that one of the largest commercial DNA testing organizations, Ancestry, is located in Utah and was founded by graduates of Brigham Young University.²⁷ For the DNA organizations themselves, the increase in law enforcement interest has required that they make critical decisions about access and searching of their sites, as well as terms of service for users. At the time of the GSK case, GEDMatch had a policy that allowed site users to opt-in to law enforcement searches.²⁸ After the GSK case the organization changed its terms of service to state, “We may disclose your Raw Data, personal information, and/or Genealogy Data if it is necessary to comply with a legal obligation such as a subpoena or warrant. We will attempt to alert you to this disclosure ...unless notification is prohibited under law.” In 2019, GEDMatch was acquired by a forensic genomics firm that has stated that it cooperates with law enforcement.²⁹

The use by law enforcement of DNA databases without permission is already more than alarming. Advances in AI technology for use in analyzing the data stored in these systems is even more disturbing. Recently, for example, researchers have touted how advances in algorithmic technology could revolutionize DNA analysis for criminal investigations.³⁰ Lacking, however, is

25. See Megan Molteni, *The Creepy Genetics Behind the Golden State Killer Case*, WIRED (Apr. 27, 2018, 2:00 PM), <https://www.wired.com/story/detectives-cracked-the-golden-state-killer-case-using-genetics/>; Heather Murphy, *She Helped Crack the Golden State Killer Case. Here's What She's Going to Do Next.*, N.Y. TIMES (Aug. 29, 2018), <https://www.nytimes.com/2018/08/29/science/barbara-rae-venter-gsk.html>; Sarah Zhang, *How a Genealogy Website Led to the Alleged Golden State Killer*, ATLANTIC (Apr. 27, 2018), <https://www.theatlantic.com/science/archive/2018/04/golden-state-killer-east-area-rapist-dna-genealogy/559070/>.

26. See Emma Coleman, *One State May Become the First to Ban Law Enforcement Use of Genealogy Databases*, ROUTE FIFTY (Jan. 21, 2020), <https://www.routefifty.com/public-safety/2020/01/utah-dna-databases/162544/>.

27. See Jennifer Graham, *The Company That Analyzed Your DNA Just Sold the Results to Someone Else. Really, What Are the Risks?*, DESERET NEWS (Aug. 21, 2018, 2:26 PM), <https://www.deseret.com/2018/8/21/20651592/the-company-that-analyzed-your-dna-just-sold-the-results-to-someone-else-really-what-are-the-risks>; Stuart Leavenworth, *DNA for Sale: Ancestry Wants Your Spit, Your DNA and Your Trust. Should You Give Them All 3?*, TAMPA BAY TIMES (June 3, 2018), https://tampabay.com/news/business/DNA-for-Sale-Ancestry-wants-your-spit-your-DNA-and-your-trust-Should-you-give-them-all-3-_168819151/.

28. See Nila Bala, *We're Entering a New Phase in Law Enforcement's Use of Consumer Genetic Data*, SLATE (Dec. 19, 2019, 7:30 AM), <https://slate.com/technology/2019/12/gedmatch-verogen-genetic-genealogy-law-enforcement.html>; Natalie Ram, *The Genealogy Site That Helped Catch the Golden State Killer Is Grappling With Privacy*, SLATE (May 29, 2019, 7:30 AM), <https://slate.com/technology/2019/05/gedmatch-dna-privacy-update-law-enforcement-genetic-genealogy-searches.html>.

29. Bala, *supra* note 28.

30. See Karen Richmond, *AI Could Revolutionise DNA Evidence – But Right Now We Can't Trust the Machines*, THE CONVERSATION (Jan. 29, 2020, 6:35 AM), <http://theconversation.com/ai->

a full consideration of what moving forward with such innovations could mean for civil rights, especially if mistakes are made in the data collection and aggregation processes.³¹ Instead, the focus has been on technology and how advances in technology might assist with data analysis. This may be a result of continued property-based language instead of language that more accurately reflects the nature of this kind of data.

Scholars have argued for other definitions of data that are useful for examining. One such definition is proposed by Ferryman under her conceptualization of data as gift as a framework for relationships related to data and participation.³² Ferryman uses Marcel Mauss' work on indigenous cultures and gift-giving as a framework for her definition. At the most basic level, Mauss found three elements for gifts: gift-giving, gift-receiving, and an obligation to reciprocate a gift.³³ From this Ferryman distills one principle: "there is no such thing as a free gift." Therefore, data should be thought of as action that comes with the obligation for the individual or organization receiving it to reciprocate in some way. In the case of health and medical research, where data collection is integral to advances in treatment and diagnosis, it would place an obligation on the researchers to provide some kind of tangible benefit to the individuals, perhaps, in the form of relationships, making communities—particularly marginalized and vulnerable groups—into stakeholders instead of mere data subjects.

While the Ferryman's concept of data as gift is meritorious and an important idea for building frameworks for interactions with communities, particularly in the public health context, it does not provide a true description of the *thing* at issue. Noted library and information scholar Christine Borgman defined data as: "Representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship."³⁴ Borgman uses this concept of data in connection to an explication of data stewardship, particularly the ideal of FAIR: findable, accessible, interoperable, and reusable. Inherent in this definition of data is the emphasis on use for research or scholarship. If using a broad definition of research, this may work. However, this definition seems limited to a particular kind of use of data. It also misses the networked nature of data.

could-revolutionise-dna-evidence-but-right-now-we-cant-trust-the-machines-129927; Chris Baraniuk, *The New Weapon in the Fight Against Crime*, BBC (Mar. 3, 2019), <https://www.bbc.com/future/article/20190228-how-ai-is-helping-to-fight-crime>.

31. See Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 15, 42 (2019), for a discussion of issues with dirty data.

32. Kadja Ferryman, *Reframing Data as a Gift* (Apr. 17, 2017) (unpublished remarks), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3000631.

33. MARCEL MAUSS, *THE GIFT: THE FORM AND REASON FOR EXCHANGE IN ARCHAIC SOCIETIES* (2002).

34. Christine L. Borgman, presentation to the National Press Club: Unstable in Concept and Context (Nov. 15, 2019), <https://escholarship.org/uc/item/0zf478ch>.

A broader and more complete definition of data is that of a *networked representation* or *observation*. This definition recognizes that data is not singular, but always comes attached with labels, contexts, and biases fastened from its inception, if not collection, and that attachments increase depending on its place in an ecosystem. An illustration from basic chemistry may be helpful. As noted above, data is usually conceptualized as a singular object, divorced from others. In that way, data is analogized to a solo atom. Yet, many of the elements on the periodic table of elements never appear as a single, solitary atom, but as molecules of more than one atom of the same nature.³⁵

A similar thing happens with data—there’s never just one datum collected, but several kinds of data. Even if a specific “data point” were examined, that particular point, too, would be endowed with other data that shape it, including the researcher’s choice in topic, participants, and research questions, among other things.³⁶ Likewise, an atom is made up of smaller molecular particles—protons, neutrons, and electrons—that shape its characteristics. An even more accurate definition of data, then, would be *a system of networked representations* or *observations*. This definition recognizes how data are used to make inferences, for evaluation, measurement, assessment, etc. of individuals, organizations, and programs. To do this, the data that represent must be arranged to decide or define relationships.

III. WHY DATA GOVERNANCE?

A critique of data governance first requires an adequate definition of governance. Governance is more than government; it “refers to all processes of governing, whether undertaken by a government, market, or network, whether over a family, tribe, formal or informal organization, or territory, and whether through laws, norms, power, or language.”³⁷ Governance can be also described as “the mechanisms, processes and institutions, through which citizens and groups articulate their interests, exercise their legal rights, meet their obligations and mediate their differences.”³⁸ It is the word “process,” including social processes, that is of great importance, signifying that governance is more than just law; governance embodies both process and structure.³⁹

More than just the process, it is the process of decision-making that is the crux of governance. Decision-making requires the recognition and use of relationships, including the connections between actions and outcomes, as well

35. DAVID W. BALL & JESSIE A. KEY, *Molecules and Chemical Nomenclature*, in *INTRODUCTORY CHEMISTRY - 1ST CANADIAN EDITION* (2014).

36. See Richard A. Berk, *An Introduction to Sample Selection Bias in Sociological Data*, 48 *AM. SOCIO. REV.* 386 (1983); Jelke Bethlehem, *Selection Bias in Web Surveys*, 78 *INT’L STAT. REV.* 161 (2010); M. Delgado-Rodríguez & J. Llorca, *Bias*, 58 *J. EPIDEMIOL. CMTY. HEALTH* 635 (2004).

37. MARK BEVIR, *GOVERNANCE: A VERY SHORT INTRODUCTION* 1 (2012).

38. UNDESA, UNDP & UNESCO, *UN SYSTEM TASK TEAM ON THE POST-2015 UN DEVELOPMENT AGENDA: GOVERNANCE AND DEVELOPMENT* (May 2012), https://www.un.org/millenniumgoals/pdf/Think%20Pieces/7_governance.pdf.

39. Peter Bogason & Juliet A. Musso, *The Democratic Prospects of Network Governance*, 36 *AM. REV. PUB. ADMIN.* 3 (2006).

as between products and services and organizations; it shapes the “rights, rules, preferences and resources that structure political outcomes.”⁴⁰ A failure in governance is the use of law as a proxy for good decision-making, in place of recognizing the impacts of law on individuals and organizations. Therefore, data governance includes interventions aimed at “chang[ing..] data-related incentives, knowledge, institutions, decision-making, and behaviors.”⁴¹ More specifically, this definition encompasses the processes, decisions, and rules that organizations—governmental, civil society, and corporate—undertake when dealing with data. This also includes any partnership or so-called co-governance agreements between different organizations or networks of organizations.

Key to data governance is an understanding of what Sean McDonald calls the digital “supply chain”—“real-time network[s]” where organizations collect and move data through a system.⁴² This means identifying the various organizations, motivations, and uses for data, as well as the possible conflicts and impact that may arise. The supply chain is integral to the data lifecycle, the stages of data from capture through interpretation and including storage. It is also important for understanding the possible impacts of data use. A recent report of the British Royal Society focused on good data governance across the lifecycle and identified issues that may arise with data that require governance infrastructure be set in place: data integrity, bias in data, accidental collection, crossing sectors, statistical profiling and stereotyping, transparency, accountability, and impact.⁴³ Current data governance regimes fail to meet both the requirements of good governance and to anticipate the issues that will emerge during the data lifecycle.

In the years since the coining of the phrase “big data,” there have been many guidelines, expressions of policy, and attempts at legislation to impede and/or mitigate the harms proffered by the collection of massive amounts of personal data. Technological advancements and the widespread centralization of personal data shared in networked platforms, too, have enticed legislators to attempt to do something about the possible harms. In the United States this has translated to the proposal and sometimes passage of bills that turn out to be sectoral, vague, underinclusive, impractical once passed, and/or so flexible so as to not really make much difference or cause added headaches for the people they were aimed at protecting.

While data governance includes legislation, failures in legislative execution demonstrate that laws are only part of the governance picture. Some attempts at data governance use a hybrid model—mixing several governance

40. JAMES G MARCH & JOHAN P OLSEN, *DEMOCRATIC GOVERNANCE* (1995).

41. Maria Carmen Lemos & Arun Agrawal, *Environmental Governance*, 31 *ANN. REV. ENV'T. RES.* 297, 298 (2006).

42. Sean Martin McDonald, *From Space to Supply Chain: Humanitarian Data Governance* (Aug. 12, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3436179.

43. THE ROYAL SOCIETY, *CONNECTING DEBATES ON THE GOVERNANCE OF DATA AND ITS USES* (July 16, 2016), http://www.webscience.org/wp-content/uploads/sites/117/2016/12/DES4610_Data-Governance-report.pdf.

schemes in an attempt to achieve policy goals. An example of a kind of hybrid model and arguably ineffective governance scheme is used in the United States in the regulation of privacy and data protection, which sees an uncoordinated mix of state and federal agencies like the Federal Trade Commission and state attorneys general, as well as state and federal legislation and regulation. Tasked with protecting consumers from unfair and deceptive business claims and practices, the FTC's power derives from federal legislation.⁴⁴ While not specifically mandated to work in the area of consumer privacy, this has come under its powers. The FTC has been involved with requiring business transparency about privacy with privacy policy regulations for example. At the same time, and under the same privacy threats, the FTC may not take on all data practices that evoke individual privacy; the Agency would need a significant increase in funding to do so. Further, under many of the privacy "regulations," individuals have no privacy right of action but must wait on the FTC to enforce the law's prohibitions. This is a concentration of power in a particular agency. This does not mean that individuals have no recourse against organizations. Of course, state and other laws may allow civil suits or state attorneys general to pursue criminal penalties. It does demonstrate how hybrid systems, without coordination of organizations and some form of omnibus structure and decision-making and foundation, can limit data protections.

Data protection is a significant concern in emerging scholarship and policy on platform governance. Although not the focus of this article, it is important to mention platform governance, which has emerged in response to various scandals and revelations about media and technology that have become integral and ubiquitous. Platform governance is a frame that recognizes that "platforms are fundamentally political actors that make important political decisions while engineering what has become the global infrastructure of free expression,"⁴⁵ while at the same time recognizing that organizations are subject to external governance. This then requires the identification of the various actors involved in platform governance including the usual suspects of government, users, and the platforms themselves, but also including related organizations like advertisers, data-brokers, and "other parties that participate in the platform's ecosystem."⁴⁶

In fact, "platform-driven ecosystems" that allow multiple actors to participate have been called the "future of the digital age."⁴⁷ Platforms—organizations that "leverage networked technologies to facilitate economic exchange, transfer information, connect people, and make predictions,"⁴⁸ continue to emerge as the business model of choice for organizations across

44. Woodrow Hartzog & Daniel J. Solove, *The Scope and Potential of FTC Data Protection*, 83 GEO. WASH. L. REV. 2230 (2015); Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014).

45. Robert Gorwa, *What is Platform Governance?*, 22 INFO. COMM'N SOC'Y 854, 857 (2019).

46. *Id.*

47. Mark Fenwick et al., *The End of 'Corporate' Governance: Hello 'Platform' Governance*, 20 EUR. BUS. ORG. L. REV. 171, 177 (2019).

48. *Id.* at 171.

several industries. Platforms and the ecosystems that emerge surrounding them, are, in fact governance networks that take, for the most part, the form of a hybrid governance network: having the participants in the network engaged with a central organization (the platform organization itself). Unlike the lead-organization governed network identified by Provan and Kenis,⁴⁹ in platform ecosystems power is concentrated in the lead organization, which governs⁵⁰ through various agreements, policies, and contracts with other actors in the system. Platforms are significant for data collection, use, security, etc. It is important, then, to comprehend the roles they play and their relationships to data and other actors and actants in the data ecosystem.

IV. AN ECOLOGICAL APPROACH TO DATA AND GOVERNANCE

Platforms provide infrastructure for parts of the data ecosystem. The study of ecology, usually considered under the umbrella of biology or biological science, is the study of systems and structures. That is, the field of ecology concerns itself with not only a specific item, say a human or badger. Instead, ecology examines the relationships between the item and other items and systems in its physical environment.⁵¹ Human ecology, in particular, centers humans as the organism of interest, with human ecology theory finding the significance in studying both the human and their social interactions.⁵²

In human development, ecological systems theory details how a person's immediate environment along with "social context, both formal and informal" in which the environments rested influenced the process of human development.⁵³ Bronfenbrenner proposed a change in the traditional method of considering human development, which focused either on naturalistic observations of humans at particular points of development, often considering only one "being" at a time and in one setting. He argued that true understanding of human development "requires examination of multiperson systems of interaction not limited to a single setting and must take into account aspects of the environment beyond the immediate situation containing the subject,"⁵⁴ which required envisioning the "environment" for a human as a model of four nested systems: micro, meso, exo, and macro. In brief, the microsystem

49. Keith G. Provan & Patrick Kenis, *Modes of Network Governance: Structure, Management, and Effectiveness*, 18 J. PUB. ADMIN. RSCH. THEORY 229 (2007).

50. Kate Klonick, *Does Facebook's Oversight Board Finally Solve the Problem of Online Speech?*, in MODELS FOR PLATFORM GOVERNANCE 51–53 (2019).

51. Sean Esbjörn-Hargens & Michael E. Zimmerman, *An Overview of Integral Ecology* 9 (Integral Institute, Resour. Pap. No. 2, 2009); David B. Lindenmayer et al., *Value of Long-term Ecological Studies*, 37 AUSTRAL ECOLOGY 745 (2012).

52. Margaret M. Bubolz & M. Suzanne Sontag, *Human Ecology Theory*, in SOURCEBOOK OF FAMILY THEORIES AND METHODS: A CONTEXTUAL APPROACH 419–50 (Pauline Boss et al. eds., 1993), https://doi.org/10.1007/978-0-387-85764-0_17.

53. Urie Bronfenbrenner, *Toward an Experimental Ecology of Human Development.*, 32 AM. PSYCH. 513 (1977).

54. *Id.* at 514.

includes the direct subject of study. In human development, this would be the human. The microsystem rests within the mesosystem, which contains that human's relationships or connections. The exosystem, which encompasses the mesosystem, includes all of the formal and informal structures that influence human development. Finally, the macrosystem represents the various environments in which the human, their relationships, and structures inhabit, including the social, political, economic, and legal, among others. According to Bronfenbrenner, this kind of ecological model represents the complexity of human development and ecology, taking into account the various things that shape who a person is and becomes.⁵⁵

This kind of ecological thinking—considering relationships and connections between things—has been used outside of the physical sciences in social sciences like psychology, mass communication, education, and social work. In the scholarship on community health interventions in particular, the approach is to consider several layers of systems to understand and shape specific outcomes for those living within a community. Stokols chronicles the social ecological approach used for studying community health campaigns.⁵⁶ This set of principles offers a “framework for understanding the dynamic interplay among persons, groups, and their sociophysical milieus.”⁵⁷ Within this context, the paradigm recognizes the physical, social, and cultural dimensions to health, and incorporates terminology from ecology such as interdependence, negative feedback, and amplification, among others. Ultimately, this framework is interdisciplinary.

Likewise, in communication research, scholars have taken ecological approaches to the study of journalism and other media structures and their influence on humans and human behavior. An example of mass communication research using an ecological approach to studying media is communication infrastructure theory. Promulgated by Sandra Ball-Rokeach and several of her colleagues, communication infrastructure theory deems storytelling networks as essential for the development and sustainment of civic engagement.⁵⁸ In particular, the theory argues that communication resources enable individuals to engage in collective action, and that “storytelling networks” set in a communication action context describes the nexus of interpersonal, organizational, and institutional communication relationships that assist in

55. *Id.* at 515.

56. D. S. Stokols, *Translating Social Ecological Theory Into Guidelines for Community Health Promotion*, 10 AM. J. HEALTH PROMOTION 282 (1996), <https://escholarship.org/uc/item/2bv79313>.

57. *Id.* at 283.

58. See Garrett M. Broad et al., *Understanding Communication Ecologies to Bridge Communication Research and Community Action*, 41 J. APPLIED COMMUN. RSCH. 325, 325–45 (2013); see also Yong-Chan Kim & Sandra J. Ball-Rokeach, *Community Storytelling Network, Neighborhood Context, and Civic Engagement: A Multilevel Approach*, 32 HUM. COMMUN. RSCH. 411, 411–39 (2006); Yong-Chan Kim & Sandra J. Ball-Rokeach, *Civic Engagement from a Communication Infrastructure Perspective*, 16 COMMUN. THEORY 173, 173–97 (2006); Holley A. Wilkin et al., *Applications of Communication Infrastructure Theory*, 25 HEALTH COMMUN. 611, 611–12 (2010).

cultivating neighborhood belonging, which can lead to civic engagement. CIT, then, requires a consideration of three different kinds of storytelling agents: micro—the residents of a community; meso—the specific neighborhood; and macro—the entire community.⁵⁹ The aim of using CIT as a framework is to assist with understanding the motivations for specific kinds of civic engagement and participation, as well as identifying systems and structures that influence participation. Of course, an environmental or ecology-related approach is not new for law and policy particularly as it relates to information. In discussions of policy and the public domain, scholars have considered the analogies of environmentalism⁶⁰ and the use of raw materials.⁶¹

An ecological approach for considering how data should be governed is appropriate because it assists with identifying the specific *thing* to be governed, that *thing*'s relationships/connections that can and/or should influence govern choices, the institutions and societal structures that impact govern and who will be tasked with enforcement and implementation, and the environment(s) in which data governance must occur. All of these many factors must be examined to achieve anywhere near a comprehensive and adequate response to the massive volume of data collection, continued surveillance, and data misuse. The next section details the four nested systems of data governance ecology in this approach, providing descriptions of ongoing genetic databases conflicts to help illustrate the importance of considering the various levels of the model.

A. *Microsystem–data representations*

Like with Bronfenbrenner's original explication of the ecological approach to human development, this ecological approach to data governance begins with the microsystem, encompassing data, defined above as networked representations. This means that within this system is the foundation layer of the data itself. Understanding the microsystem essential for good ecological governance because it identifies the two functions of data: as thing and as action. Data as thing applies Buckland's conceptualization of "information as thing"—"objects . . . that are regarded as being informative"⁶² to data, making it of interest to research systems. Data as a thing allows organizations to make inferences. As the volume of data increases, uncertainty and equivocality decrease for organizations. Therefore, data *as thing* can be viewed as evidence, "though without implying that [the evidence is] necessarily accurate, useful, or even pertinent to the user's purposes."⁶³ Important in this idea of data-as-thing-

59. See Kim & Ball-Rokeach, *supra* note 58.

60. See James Boyle, *A Politics of Intellectual Property: Environmentalism for the Net?*, 47 DUKE L.J. 87 (1998); Anupam Chander & Madhavi Sunder, *The Romance of the Public Domain*, 92 CALIF. L. REV. 1331, 1331–74 (2004).

61. See Julie E. Cohen, *The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy*, 31 PHIL. TECH. 213, 213–33 (2018).

62. Michael K. Buckland, *Information as Thing*, 42 J. AM. SOC'Y INFO. SCI. 351, 351 (1991).

63. *Id.* at 353.

as-evidence is the implication that the “thing” is passive – it does nothing, but something is done to it or with it.

In contrast, data as action data as process; it changes what organizations know; it informs. Data as action also harkens to Ferryman’s definition of data as requiring reciprocity. In the case of organizations collecting data, this will mean that the act of data collection initiates certain duties.⁶⁴ Future research must consider both data as thing and data as action to inform about how they behave in the data ecosystem, and their relations.

B. Meso–data relationships

Surrounding the microsystem is the mesosystem, which considers the relationships and connections to data. If our definition of data is that of a networked representation. It is important to consider what things are in the network. Bronfenbrenner describes the mesosystem as a system of microsystems. For data, this would mean examining both the connections that data has with other kinds of data as well as the attachments and labels connected with the data. The mesosystem also encompasses all of the structures or settings that shape data over the life cycle.

In the GEDMatch DNA case in which the Golden State Killer was identified through law enforcement use of a commercial DNA database, understanding the nature of the various relationships that were implicated in a DNA search may have at least given state legislators pause about the kinds of laws necessary to ensure user privacy. On the federal level, of course, the Genetic Information Non-discrimination Act (GINA) exists to prohibit discrimination based on genetic data. But GINA is narrowly focused on discrimination and only applies to the healthcare and employment sectors. Several state laws focused on genetic non-discrimination also exist; these too narrowly focused on insurance and employment. This fails to prevent use of genetic data for purposed beyond particular expectations and ignores the relational harms that can be caused by misuse of the data. Future research, then, must identify and investigate the influence of these data relationships, in order to create adequate policy for governance.

C. Exo–institutions/community

The exosystem includes all of the societal and institutional structures that mediate data and data governance. These structures will be tasked with implementing and enforcing data governance, but also implicate how they use data, which leads to the need for stronger data governance regulation. This system also identifies the kinds of infrastructure necessary for providing adequate data governance and embodies all of the formal and informal social structures that influence the data lifecycle. This includes all of the major societal institutions, technology, and platforms, as well as governmental and

64. See Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183–34 (2016); Ariel Dobkin, *Information Fiduciaries in Practice: Data Privacy and User Expectations*, 33 BERKELEY TECH. L. J. 1, 1–50 (2018); Jonathan Zittrain, *Engineering an Election*, 127 Harv. L. Rev. F. 335, 335–41 (2014).

civil society organizations.⁶⁵ That data encounters over the lifecycle or that shapes how data moves through the lifecycle.

At the same time, an exosystem “has been defined as consisting of one or more settings that do not involve ... an active participant but in which events occur that affect, or are affected by, what happens in that setting.”⁶⁶ Therefore, exosystems concern settings and actors that indirectly affect data. It is important to investigate the structures in the exosystem as though they may not have direct connections to individuals, they nonetheless may impact how data is collected, used, etc. As an example, many social media sites, employment databases for various public professions, as well as law enforcement make photos available online. These photos have become the fuel for organizations developing facial recognition systems that scrape social media and other databases to train their software. An ecological approach would examine these facial recognition platforms, their attending organizations, as well as the spaces where they collect data. It will also be important to understand the law, or lack thereof, as boundary infrastructure in this line of research.

D. Macro-environment(s) i.e., social, political, economic, etc.

The macrosystem is the environment(s) in which the micro, meso, and exo systems rest. Under Bronfenbrenner’s conceptualization, the macrosystem “refers not to specific contexts ... but to general prototypes, existing in the culture or subculture, that set the pattern for the structures and activities occurring at the concrete level.”⁶⁷ These prototypes are called the “blueprints” for society because they hold true for both formal and informal settings.⁶⁸ Culture is expressed in law, customs, belief systems, economic structures, etc. It is important for future research to examine how these prototypes shape all of the other systems. For data, this would mean thorough investigations of how culture and subcultures shape the environment for data and are then expressed in how various organizations and individuals respond.

At the same time, it is important to consider how culture actually behaves. A criticism of Bronfenbrenner’s theory is that it relegates culture to the macrosystem, as though culture does not permeate all systems in human ecology.⁶⁹ Culture permeates *everything*, structure, and setting in society. Therefore, ignoring culture presents a view of data and the institutions and organizations connected with data as neutral, in spite of the overwhelming evidence to the contrary.⁷⁰ This requires a reconsideration of how the macrosystem operates. While traditionally viewed as the most outer oval of the

65. See Bronfenbrenner, *supra* note 53, at 515.

66. URIE BRONFENBRENNER, THE ECOLOGY OF HUMAN DEVELOPMENT 237 (1979).

67. See Bronfenbrenner, *supra* note 53, at 515.

68. See Nicole M. Vélez-Agosto et al., *Bronfenbrenner’s Bioecological Theory Revision: Moving Culture from the Macro into the Micro*, 12 PERSPS. PSYCH. SCI. 900, 902 (2017).

69. See *id.* at 906.

70. See EUBANKS, *supra* note 3; SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); see generally O’NEIL, *supra* note 19.

model, or the largest nesting doll, it would be more accurate to view the macrosystem as closer to atmospheric, flowing through all other systems and circulating throughout the various systems, closely related to Appadurai's conception of scapes that influence information flows and denote the fluidity of five dimensions of cultural flows.⁷¹ Though Appadurai applied this suffix to characteristics in relation to the international capital, it works with data, which too is subject to international flows.

V. DATA ECOLOGY IN ACTION

Ecology recognizes that things within a system, or a system of systems, impact and/or change other things. But ecology and ecosystems are not clean models; they are messy.

This is an actual ecosystem – containing representations (data) that encounter other data, and communities, where institutions reside that enforce and enable flows, all embedded within economic, legal, and social, among other, environments. The messiness of this system is the point. A simple analogy, like that of property, does not work for a system like data where humans are involved.

A recent controversy with the U.S. National Institutes of Health All of Us precision medicine initiative illustrates the necessity of considering an ecological approach to data governance. All of Us is a research program that aims to recruit one million people in the U.S. from which to gather health data and specimens.⁷² This information will be used for biomedical research and includes “health questionnaires, electronic health records (EHRs), physical measurements, the use of digital health technology, and the collection and analysis of biospecimens.”⁷³ The system is also billed as allowing researchers “researchers to take into account individual differences in lifestyle, socioeconomic factors, environment, and biologic characteristics in order to advance precision diagnosis, prevention, and treatment.”⁷⁴

But the breadth of this research data and the possible inferences that can be made from it are of concern for several tribal communities. In the U.S., there are nearly 600 federally recognized Tribal governments, which exercise sovereignty over many vectors of tribal life including public health data.⁷⁵ In 2018, however, it was reported that the NIH was bypassing tribal data sovereignty to collect the data by recruiting in urban areas containing large populations of Native Americans, without consulting with tribes or the National

71. See Arjun Appadurai, *Disjuncture and Difference in the Global Cultural Economy*, 7 THEORY CULTURE & SOC'Y 295, 295–310 (1990).

72. See *The “All of Us” Research Program*, 381 NEW ENG. J. MED. 668, 668 (2019); see also National Institutes of Health (NIH)—All of Us, NAT'L INST. OF HEALTH (NIH) (2020), <https://allofus.nih.gov/future-health-begins-all-us> (last visited Feb 20, 2021).

73. *The “All of Us” Research Program*, *supra* note 72.

74. *Id.*

75. See Aila Hoss, *Exploring Legal Issues in Tribal Public Health Data and Surveillance*, 44 S. ILL. U. L.J. 27, 27 (2019).

Congress of American Indians.⁷⁶ At issue is the sharing of EHRs and other data with pharmaceutical and other organizations. Further, a question remains about the applicability of the Health Insurance Portability and Accountability Act (“HIPAA”) to the organizations would be able to access the information.⁷⁷

For some, the actions of the NIH and other researchers and programs who have sought indigenous data is a form of biocolonialism—the assertion of control, ownership, and use of biological data and specimens without or beyond the guidance of tribal governments and without direct benefit.⁷⁸ The result has been a call for both decolonizing data⁷⁹ and for the recognition of indigenous data sovereignty.⁸⁰ The move for indigenous data sovereignty has been long, but the first recognized formal international convening happened in 2015, with a meeting of indigenous researchers in Australia.⁸¹ Following this, collectives of indigenous formed groups and established charters aimed at creating guidance for data sovereignty. In the U.S. one such group is the US Indigenous Data Sovereignty Network (“USIDSN”) that aims to “promot[e] Indigenous data sovereignty through decolonizing data and Indigenous data governance.”⁸² For collectives like USIDSN the principles of data sovereignty reflect a different framework than that traditionally used in the Western governance. Sovereignty, under tribal governance may take several forms,⁸³ but offers a way forward for tribes with the aim of protecting privacy, preempting extractive

76. See Terri Hansen & Jacqueline Keeler, *The NIH Is Bypassing Tribal Sovereignty to Harvest Genetic Data from Native Americans*, VICE, Dec. 21, 2018, <https://www.vice.com/en/article/8xp33a/the-nih-is-bypassing-tribal-sovereignty-to-harvest-genetic-data-from-native-americans>; Kalen Goodluck, *Indigenous Data Sovereignty Shakes Up Research*, HIGH COUNTRY NEWS, Oct. 8, 2020, <https://www.hcn.org/issues/52.11/indigenous-affairs-covid19-indigenous-data-sovereignty-shakes-up-research>.

77. See Hansen & Keeler, *supra* note 76.

78. See Manola Secaira, *Abigail Echo-Hawk on the Art and Science of ‘Decolonizing Data,’* CROSSCUT, May 31, 2019, <https://crosscut.com/2019/05/abigail-echo-hawk-art-and-science-decolonizing-data>; see also Hansen & Keeler, *supra* note 76; Kalen Goodluck, *Covid is Strengthening the Push for Indigenous Data Control*, WIRED, Oct. 10, 2020, <https://www.wired.com/story/covid-is-strengthening-the-push-for-indigenous-data-control/>.

79. See Secaira, *supra* note 78.

80. See Goodluck, *supra* note 76.

81. See *Te Mana Raraunga – Māori Data Sovereignty Network Charter* (2015), MĀORI DATA SOVEREIGNTY NETWORK, <https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5913020d15cf7dde1df34482/1494417935052/Te+Mana+Raraunga+Charter+%28Final+%26+Approved%29.pdf> (last visited Feb. 20, 2021).

82. United States Indigenous Data Sovereignty Network, About, <https://web.archive.org/web/20220423084152/https://usindigenousdata.org/about-us> (last visited Feb. 20, 2021).

83. See Krystal S. Tsosie, *Models of Data Governance and Advancing Indigenous Genomic Data Sovereignty*, in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 3592* (2020), <https://doi.org/10.1145/3394486.3411072> (last visited Feb. 20, 2021).

research, and recognizing the implications of data use on the many interconnected facets of life for Native Americans.⁸⁴

The All of Us controversy demonstrates both the need for more adequate data governance that recognizes the implications of the data ecosystem, as well as the need for action ensuring that the awareness of these systems is included in the development of its frameworks. Good governance, in general, is collective, responsive, equitable, and lawful. In studying and enacting ecological data governance, we must use collective and participatory approaches to the creation of frameworks. This requires engagement with traditionally marginalized and vulnerable communities, many of whom are disparately impacted by data collection and uses. It also demands that organizations—whether civic, civil society, or corporate—be responsive to collective governance decisions. Accountability necessitates legislation as an encouragement. Legislation also acts as infrastructure, and good data governance requires infrastructure, which will include platforms and mechanisms that perform the frameworks produced.

This essay has sought to provide a brief overview of a way forward for considering and governing the materials that feed the ever-burgeoning AI technological ecosystem. It further provides a research agenda for exploring exactly how this framework could work while focusing on the various aspects of the data governance scheme. It will be ever more important to investigate methods of harm reduction as the use of algorithmic systems expands.

84. See Rebecca Tsosie, *Tribal Data Governance and Informational Privacy: Constructing "Indigenous Data Sovereignty,"* 80 MONT. L. REV. 230, 231 (2019); See also Hoss, *supra* note 75, at 28.