Barbarians at the Gate: Consumer-Driven Health Data Commons and the Transformation of Citizen Science

Barbara J. Evans[1]

Abstract. This article introduces consumer-driven health information commons, which are institutional arrangements to empower groups of consenting individuals to collaborate to assemble powerful, large-scale health data resources for use in scientific research, on terms the group members themselves would set. Many of the grand scientific challenges of the 21st-century require access to big data resources that are both deeply descriptive (in the sense of containing detailed personal health, genomic, environmental, and lifestyle information about each individual) and inclusive (in the sense of supplying such data for many—sometimes, hundreds of millions—of individuals). These grand challenges include President Obama's Precision Medicine and Brain Initiatives, the Vice President's Cancer Moonshot, efforts by the National Institutes of Health and the Food and Drug Administration to clarify the clinical significance of genetic variants and to make modern diagnostics safe and effective, and attempts to create a "learning health care system" that harnesses healthcare data to improve patient care. Without big data resources, these efforts will fail.

Current research and privacy regulations, which were designed for clinical research and for small-data studies of the past, cannot support creation of the vast data resources that 21st-century science needs. These regulations enshrine data-holders (hospitals, insurers, and other entities that store people's data) as the prime movers in assembling large-scale data resources for scientific use and rely on mechanisms—such as de-identification of data and waivers of individual consent—that are unworkable going forward. They shower individuals with unwanted, paternalistic protections—such as barriers to access to their own research results—while denying them a voice in what will be done with their data.

The barbarians—the people whose data scientific researchers wish to study—are at the gate demanding our right of democratic data self-governance. The atomistic vision of individual autonomy enshrined in 20th-century bioethics ultimately disempowered the very patients and research subjects it sought to protect, empowering them to make decisions as individuals, but only as individuals, and lacking a roadmap for collective action. Individuals acting alone are strong, but never as strong as individuals acting together. This article draws on the natural resource commons theory associated with Elinor Ostrom to propose an alternative approach that places consumers—patients, research subjects, and persons who track their health using mobile and wearable sensor devices—at the center of efforts to assemble large-scale data resources of the people, by the people, and for the people, who would themselves control the terms of use through collective self-governance processes.

---

[1] This article is forthcoming in *American Journal of Law and Medicine*, Vol. 42, Issue 4 (2016). Posted with permission of *American Journal of Law and Medicine*. The author is Professor of Law, George Butler Research Professor, and Director, Center for Biotechnology & Law, University of Houston Law Center, bjevans@central.uh.edu. This is a commissioned paper of the Robert Wood Johnson Foundation's Health Data Exploration Project (Kevin Patrick, M.D., M.S., PI) http://hdexplore.calit2.net. Portions of the analysis draw on research funded by NIH/NHGRI grants U01HG006507 (GPJ) and U01HG007307-02S2 (GPJ). The author thanks Cinnamon Bloss, Matthew Bietz, and Kevin Patrick for their helpful insights.

Barbarians at the Gate: Consumer-Driven Health Data Commons and the Transformation of Citizen Science

> *"The expression 'barbarians at the gate' was … used by the Romans to describe foreign attacks against their empire [or] in contemporary English within a sarcastic, or ironic context, when speaking about a perceived threat from a rival group of people, often deemed to be less capable, or somehow 'primitive.'"*[2]

Introduction

Citizen science has a mixed reputation. It includes well-organized, crowd-sourced efforts like the 1830s British great tidal experiment that enlisted members of the public to monitor tides at 650 coastal locations, but it also includes less rigorous dabbling by amateurs.[3] Successful citizen science projects often engage laypeople, supervised by professional scientists, to collect and analyze data or even to assist in creating the finished scholarly works[4]—in other words, the citizens are a source of labor. This article explores an alternative model—citizen research sponsorship—in which citizens supply essential capital assets to support research. The assets could be monetary (research funding, for example), although this article focuses on a different kind of capital: data resources, which are a critical input[5] for 21st century biomedical research.

The citizen research sponsorship model flips the traditional control relationship of citizen science. Instead of laypeople laboring under the supervision of professional scientists, the professional scientists work at the instigation of citizen groups, using the people's data for projects the people endorse. Citizen groups that control an essential research input, such as data or biospecimens, sometimes succeed in leveraging their asset to enlist qualified scientists to generate desired knowledge. This sponsorship model was exemplified late in the 1980s when a group of Canavan-disease-affected families developed a disease registry and biospecimen bank and leveraged these resources to spur discovery of associated genetic variants and development of a diagnostic test.[6] Their sponsorship took the form of supplying data and biospecimens for the research, as opposed to providing funding—and this revealed a new dynamic in the era of informational research[7] that mines preexisting health records and data derived from biospecimens: Money will follow a good data resource, instead of data resources following (and having to be generated by) those who hold money. Data resources are a central currency of 21st-century science, and the question is, "Who will control them?"

---

2 Statement attributed to Christopher Adam, http://references-definitions.blurtit.com/76895/what-is-the-meaning-of-barbarians-at-the-gate

3 Michael J. Madison, *Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo*, 209, 215 *in* GOVERNING KNOWLEDGE COMMONS (Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandberg, eds. 2014).

4 *Id*.

5 U.S. Dept of Health and Human Servs. et al., Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,938 (Sept. 8, 2015) (characterizing a shift in research activities to include more informational studies that rely on data and biospecimens, in contrast to clinical studies).

6 Greenberg v. Miami Children's Hospital Research Institute, 264 F. Supp.2d 1064, 1066-67 (U.S. Dist. Court, S. D.F. *2003).*

7 U.S. Dep't of Health & Human Servs., *supra* note 5.

The Canavan families' scientific success was later marred by litigation when their chosen investigator elected to patent his discoveries and charge royalties on the test.[8] They had naively assumed he would put his discoveries into the public domain.[9] Citizen sponsors, like any other research sponsors, need well-drafted research agreements if they want to avoid unpleasant surprises. The Canavan families' greatest contribution to science ultimately may have been that they demonstrated the power of well-organized citizen groups—perhaps, next time, with appropriate consulting and legal support—to instigate high-quality scientific research. Hiring lawyers and scientists is relatively straightforward if a citizen group has money, and money need not always come from external fundraising and donors. A citizen group that controls a critical data resource, coupled with a workable revenue model, may be able to monetize its resource lawfully and on terms ethically acceptable to the group members.

This article defines the concept of consumer-driven data commons, which are institutional arrangements for organizing and enabling citizen research sponsorship. This model differs from existing arrangements for supplying personal health data for use in biomedical research. This article explores how traditional arrangements, imbedded in major federal research and privacy regulations, conceive institutional data holders—entities such as hospitals, research institutions, and insurers that store people's health data—as the prime movers in assembling large-scale data resources for research and public health. Consumer-driven data commons also differ from many of the patient-centered data aggregation models put forward as alternatives to letting data holders control the fate of people's data. One such alternative is a personally controlled electronic health record with granular individual consent:[10] that is, a scheme in which individuals (or their designated agents) gather and assemble their own health data and then specify, in very granular detail, the particular data uses that would be acceptable to each individual.

The consumer-driven data commons proposed here, in contrast, would aggregate data for a group of participating volunteers who, thereafter, would employ processes of collective self-governance to make decisions about how the resulting data resources—in the aggregate, as a collective data set—can be used. The group's collective decisions, once made, would be binding on all members of the group (at least until a member exited the group), but the decisions would be made by the group members themselves, according to rules and processes they established. This article explores the promise and the challenge of enabling consumer-driven data commons as a mechanism for consenting individuals to assemble large-scale data resources. Twenty-first

---

8 264 F.Supp.2d at 1067.

9 *Id*. At 1068 ("Plaintiffs allege that at no time were they informed that Defendants intended to seek a patent on the research. Nor were they told of Defendants' intentions to commercialize the fruits of the research…").

10 *See* Mark A. Hall & Kevin A. Schulman, *Ownership of Medical Information*, 301 JAMA 1282, 1283–84 (2009) (discussing advantages of patient-controlled longitudinal health records and suggesting that one way to foster the development of such records would be to "give patients the rights to sell access to their records, rights that are superior to the property rights held by [entities that currently hold patients' data]"); *see also* Mark A. Hall, *Property, Privacy, and the Pursuit of Interconnected Electronic Medical Records*, 95 IOWA L. REV. 631, 651 (2010) ("[I]f patients were given ownership of their complete medical treatment and health histories, they could license to compilers their rights to that information in a propertized form that could be more fully developed and commercialized."). *See also* Eric M. Meslin & Peter H. Schwartz, *How Bioethics Principles Can Aid Design of Electronic Health Records to Accommodate Patient Granular Control*, 30(Supp1) J GEN INTERN MED. 3-6 (Jan. 30, 2015) (discussing granular consent).

century science, as discussed below,[11] needs large-scale, deeply descriptive, and inclusive data resources. Granular, individual consent is inimical to the creation of such resources, which require collective action.

This article construes personal health data (PHD) broadly to include data about patients, research participants, and people who use sensor devices or direct-to-consumer testing services (together, "consumers"). PHD includes traditional sources of health data, such as data from consumers' encounters with the healthcare system as well as data generated when they consent to participate in clinical research. PHD also may include individually identifiable (or re-identifiable) research results that investigators derive during informational research—research that uses people's data or biospecimens with or without their consent. Increasingly, PHD includes genetic and other diagnostic information that healthy-but-curious consumers purchase directly from commercial test providers, as well as information people generate for themselves using mobile, wearable, and at-home sensing devices. More creepily, PHD also includes data captured passively by the panopticon of algorithms that silently harvest data from online shopping, professional, leisure, and social communication activities.[12] Such algorithms may support excruciatingly personal inferences about an individual's health status—for example, pregnancy—that arouse intense privacy concerns.

There are many competing visions of the public good and how to advance it. This analysis presumes, as its starting point, that the public good is served when PHD are accessible for biomedical research, public health[13] studies, regulatory science,[14] and other activities that generate knowledge to support continuous improvements in wellness and patient care. The goal here is not to debate this vision but rather to assume it and study how competing legal and institutional arrangements for data sharing may promote or hinder the public good and address people's concerns about privacy and control over their PHD.

1. The Challenge of Assembling Data Resources for Public Good

Consumer-driven data commons have the potential to elevate citizen science from its perceived status as do-it-yourself puttering and transform it into a force for addressing some of the grand scientific challenges of the 21st century. These challenges include President Obama's Precision Medicine[15] and Brain[16] Initiatives, the Vice President's Cancer Moonshot,[17] ongoing efforts to

---

11 *See* discussion *infra* part 2.

12 *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015) (discussing modern data-mining practices that generate health-related information about consumers).

13 *See* LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW 4 (2d ed. 2008) (describing public health as including population-oriented (as opposed to patient-specific) efforts "to ensure the conditions for people to be healthy" and "to pursue the highest possible level of physical and mental health in the population, consistent with the values of social justice"); Paul J. Amoroso & John P. Middaugh, *Research vs. Public Health Practice: When Does a Study Require IRB Review?*, 36 PREVENTIVE MED. 250, 250 (2003) (offering, as examples public health activities, tracking communicable diseases, investigating disease outbreaks, and collecting personal data to protect public health).

14 *See* U.S. Food and Drug Admin., Advancing Regulatory Science, http://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/default.htm?utm_campaign=Goo.

15 *See* The White House, The Precision Medicine Initiative, at https://www.whitehouse.gov/precision-medicine.

16 *See* The White House, Brain Initiative, at https://www.whitehouse.gov/share/brain-initiative.

clarify the clinical significance of genomic variants and ensure that modern diagnostics are safe and effective,[18] and efforts to develop a "learning health care system"[19] that routinely captures data from treatment settings—and from people's experiences as non-patients before and after their healthcare encounters—to glean insights to support continual improvements in wellness and patient care.

These scientific challenges all share a common feature: they require access to very large-scaled data resources—sometimes, data for tens to hundreds of millions of individuals[20] (known as "data partners"[21] in the nomenclature of the Precision Medicine Initiative). The most valuable data resources are deeply descriptive in the sense of reflecting, for each individual, a rich array of genomic and other diagnostic test results, clinical data, and other available PHD such as data from mobile and wearable health devices that may reflect lifestyle and environmental factors that influence health.[22] The data need to be longitudinal in the sense of tracing, as completely as possible, the history of a person's innate characteristics, factors that may have influenced the person's health status, diagnoses during spells of illness, treatments, and health outcomes.[23]

Such data, unfortunately, are inherently identifiable, both because they have to be, and because they can be: They *have to be*, because access to identifiers is necessary, at least in certain phases of database creation, in order to link each person's data that is arriving from different data holders, to verify that the data all pertain to the same individual, and to update the person's existing data with subsequent clinical observations.[24] Moreover, the resulting assemblage of data—deeply descriptive of each individual—potentially *can be* re-identified, even if overt identifiers like names are stripped off after the data are brought together.[25] If a

---

17 *See* The White House, Fact Sheet: Investing in the National Cancer Moonshot (Feb. 1, 2016), at https://www.whitehouse.gov/the-press-office/2016/02/01/fact-sheet-investing-national-cancer-moonshot.

18 *See* U.S. Food and Drug Admin., Public Workshop - Use of Databases for Establishing the Clinical Relevance of Human Genetic Variants, November 13, 2015 at http://www.fda.gov/medicaldevices/newsevents/workshopsconferences/ucm459450.htm; *see also*, Barbara J. Evans, Wylie Burke & Gail P. Jarvik, *FDA and Genomic Tests: Getting Regulation Right*, 372 NEW ENGLAND JOURNAL OF MEDICINE 2258-64 (2015).

19 INSTITUTE OF MEDICINE, IOM ROUNDTABLE ON EVIDENCE-BASED MEDICINE, THE LEARNING HEALTHCARE SYSTEM 3, 6 (Olsen, Aisner, McGinnis, eds.).

20 *See generally* B.A. Shirts *et al., Large Numbers of Individuals Are Required to Classify and Define Risk for Rare Variants in Known Cancer Risk Genes*, 16 GENETICS IN MEDICINE 529-34 (2014) (discussing the size of data resources required to draw inferences about the clinical significance of rare genetic variants).

21 The White House, Fact Sheet: Obama Administration Announces Key Actions to Accelerate Precision Medicine Initiative (Feb. 25, 2016), at https://www.whitehouse.gov/the-press-office/2e a 016/02/25/fact-sheet-obama-administration-announces-key-actions-accelerate.

22 Barbara J. Evans, *Much Ado About Data Ownership*, 25 HARVARD JOURNAL OF LAW & TECHNOLOGY 69 , 90 (2012), *available at* http://jolt.law.harvard.edu/articles/pdf/v25/25HarvJLTech69.pdf

23 *Id.*

24 *Id*. at 93-94.

25 *See* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: A PROPOSED FRAMEWORK FOR BUSINESSES AND POLICYMAKERS 35–38 (2010) (warning that the distinction between personally identifiable information and non-identifiable information is increasingly irrelevant in light of the potential for data to be re-identified); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV.

dataset contains a rich, multi-parametric description of somebody, there may be only one individual in the world for whom all of the parameters are a match. If other, external datasets link a subset of those parameters to the person's identity, re-identification is possible.[26]

For some important types of research, the data resources also need to be highly inclusive, in the sense that most (or even all) people are included in the dataset and share their data for research.[27] Inclusive data sets capture rare events and allow them to be studied and, more generally, they avoid consent bias (selection bias).[28] Empirical studies suggest that people who consent to having their data used in research may have different medical characteristics than the population at large.[29] For example, patients who are sick and have symptoms may feel more motivated than asymptomatic people are to volunteer for studies that explore possible genetic causes of their symptoms. If true, then a cohort of consenting research subjects may over-represent people who carry a specific gene variant and also happen to be ill. The study may reach biased conclusions misstating how often the variant results in actual illness.

Consent bias reportedly was a factor that contributed to a tendency for early studies to overstate the lifetime risk of breast and ovarian cancer in people with certain BRCA genetic mutations.[30] Costs of testing were high under the gene patenting doctrine of the day; insurance reimbursement criteria tended to make clinical BRCA testing available only to people with a personal or family history of these cancers; such people also were highly motivated to share their data for use in research. As a broader population gains access to BRCA testing, the available data resources are gradually expanding to include more people who have mutations without

---

1701, 1706 (2010) (discussing the risks to individual privacy if de-identified data were to be re-identified); Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, 10 AM. J. BIOETHICS 3, 5 (2010) ("Despite using various measures to deidentify health records, it is possible to reidentify them in a surprisingly large number of cases . . . .").

26 *See supra* note 25.

27 Evans, *supra* note 22, at 95-96.

28 *Id*.

29 *See generally* Brian Buckley et al., *Selection Bias Resulting from the Requirement for Prior Consent in Observational Research: A Community Cohort of People with Ischaemic Heart Disease*, 93 HEART 1116 (2007); David Casarett, Jason Karlawish, Elizabeth Andrews & Arthur Caplan, *Bioethical Issues in Pharmacoepidemiologic Research*, *in* PHARMACOEPIDEMIOLOGY 587, 593-94 (Brian L. Strom ed., 4th ed. 2005); COMM. ON HEALTH RESEARCH AND THE PRIVACY OF HEALTH INFO.: THE HIPAA PRIVACY RULE, INST. OF MED., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 209-14 (Sharyl J. Nass, Laura A. Levit & Lawrence O. Gostin eds., 2009) [hereinafter IOM, PRIVACY REPORT], *available at* http://www.nap.edu/catalog/12458.html (surveying studies of consent and selection bias); Khaled El Emam et al., *A Globally Optimal k-Anonymity Method for the De-identification of Health Data*, 16 J. AM. MED. INFO. ASS'N. 670, 670 (2009); Steven J. Jacobsen et al., *Potential Effect of Authorization Bias on Medical Record Research*, 74 MAYO CLINIC PROC. 330 (1999); Jack V. Tu et al., *Impracticability of Informed Consent in the Registry of the Canadian Stroke Network*, 350 NEW ENG. J. MED. 1414 (2004); Steven H. Woolf et al., *Selection Bias from Requiring Patients to Give Consent to Examine Data for Health Services Research*, 9 ARCHIVES FAM. MED. 1111 (2000).

30 Toby Bloom, Ph.D., Deputy Director, Bioinformatics, New York Genome Center, *Next Generation Sequencing and Bioinformatics: A Researcher's Perspective*, remarks at the American Association of Law Schools Annual Conference, BioLaw Section (Jan. 8, 2016), available at https://soundcloud.com/aals-2/biolaw-co-sponsored-by-law-medicine-and-health-care/s-BDCUD

developing cancer, and lifetime risk estimates are trending downward.[31] Getting these numbers right has obvious impact on future patients who face decisions based on their test results.

One possible policy for creating large, deeply descriptive, inclusive datasets free of consent bias is to force all citizens to contribute their data, requiring them to pay a "data tax" just as we all must pay income taxes. That idea is repugnant to many, and I do not propose it except to contrast it with a rarely considered policy that this article seeks to advance: Why not get people to *want* to participate in large-scale, deeply descriptive, inclusive datasets free of consent bias? Why not make participation interesting and enjoyable, perhaps even fun? Current ethical and regulatory frameworks that govern data access, such as the Common Rule[32] and Health Insurance Portability and Accountability Act of 1996 ("HIPAA") Privacy Rule[33] decisively fail to do this. Have we, as a society, unwittingly embraced a prune-faced framework of bioethics, such that making data partnership fun would be coercive or ethically problematic and, if so, how did it come to this and what can we do about it?

The Common Rule and HIPAA Privacy Rule both provide workable pathways to obtain data—if necessary, without consent—for socially important research, public health purposes, and regulatory science.[34] But it is never fun and unconsented data access, even when it is legal, will always be controversial. Surveys show that "a majority of consumers are positive about health research and, if asked in general terms, support their medical information being made available for research"[35]—in other words, they see research participation as potentially fun—but they want to be asked before their data are taken and they prefer for their data be de-identified.[36] Sadly, as just noted, de-identification may no longer be feasible, and even if it were feasible, it cannot support creation of deeply descriptive, longitudinal data that 21st-century science needs.[37]

The existing regulations, which were designed for clinical research and for small-data informational studies of the past, function well enough and may continue to function, at least for those who are sufficiently well-lawyered to thread the needle of data access. But they do not excite people about becoming partners in the grand scientific challenges of the 21st century, which ought to be easy given how fascinating these challenges are. Current regulations sometimes insult the very people whose data investigators want to use, showering individuals with unwanted, paternalistic protections—such as barriers to access to return of their own

---

31 *Id*.

32 45 C.F.R. pt. 46.

33 Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29 and 42 U.S.C.); *see* Privacy Rule at 45 C.F.R. pts. 160, 164 (2010).

34 *See* Evans, *supra* note 22, at 82-86 (discussing nonconsensual access pathways under the Privacy Rule and Common Rule).

35 *See generally*, IOM, Privacy Report, *supra* note 29. *See generally* Leonard J. Kish & Eric J. Topol, *Unpatients—why patients should own their medical data*, 11 NATURE BIOTECHNOLOGY 921-924 (2015) (discussing individuals' willingness to participate in research); Eric Topol, *The big medical data miss: challenges in establishing an open medical resource*, 16 *Nature Reviews Genetics* 253-254 (2015) (same); HEALTH DATA EXPLORATION PROJECT, PERSONAL HEALTH DATA FOR THE PUBLIC GOOD: NEW OPPORTUNITIES TO ENRICH UNDERSTANDING OF INDIVIDUAL AND POPULATION HEALTH (2014), at http://hdexplore.calit2.net/wp-content/uploads/2015/08/hdx_final_report_small.pdf (same, in the context of non-traditional forms of PHD).

36 IOM Privacy Report, *supra* note 29, at 82.

37 *See supra* notes 23-26 and accompanying text.

research results[38]—while denying them a voice in what will be done with their data. Data partners' only real "voice" is their right to withhold consent, take their data, and go home—and even that right can be waived by an Institutional Review Board,[39] typically staffed by employees of institutions that wish to use the people's data and whom the people never chose to represent their interests.[40]

Most people have no wish to take their data and go home. Surveys suggest that 80% of Americans would like to see their data used to generate socially beneficial knowledge.[41] They want to participate, but subject to privacy, data security protections, and other terms that are transparent and satisfactory to themselves. Consumer-driven data commons are a vehicle for enabling consumers to set and enforce those terms through collective self-governance and to find the voice that ethics and regulatory frameworks consistently deny them.

2. Distinguishing Data Ownership, Data Commons, and the Public Domain

There are multiple, viable pathways for developing heath data commons to promote public good, and it will be important for policymakers to have the wisdom to allow them to evolve in parallel during early phases of the effort.

- The first major pathway,[42] resembling propertization, bestows entitlements (such as specific rights of access, rights to transfer and enter transactions involving data, rights to make managerial decisions about data, or even outright data ownership) on specific parties. It then relies on those parties to enter private transactions to assemble large-scale data resources. The initial endowment of rights can be bestowed various ways: on the individuals to whom the data relate (patients and consumers); on data holders such as hospitals, insurers, research institutions, and manufacturers of medical and wearable devices that store and possess people's data; on both groups; or on other decision-makers.
- A second major pathway is to develop data resources in the public domain[43]—for example, through legislation or regulations that force entities that hold data to supply it for specific public health or regulatory uses, or by using public funds (e.g., grants, tax incentives) to create data resources under rules that make them openly available for all to use (or for use by a designated group of qualified entities, such as public health officials or biomedical researchers, who are legally authorized use data on the public's behalf).
- A third pathway is to foster creation of data commons, which are distinct from the other two pathways and can include many different types of commons that may exist simultaneously.[44]

---

38 *See generally* Barbara J. Evans, *The First Amendment Right to Speak About the Human Genome*, 16 U. PENN. JOURNAL OF CONSTITUTIONAL LAW 549 - 636 (2014) (discussing and criticizing IRB-imposed restrictions on return of results from research),

39 45 C.F.R. § 164.512(i) (2010) (HIPAA waiver provision); *id.* § 46.116(d) (Common Rule waiver provision).

40 *See* Carl H. Coleman, *Rationalizing Risk Assessment in Human Subject Research*, 46 ARIZ. L. REV. 1, 13–17 (2004) (discussing procedural informality of the Common Rule).

41 *See* Kish & Topol, *supra* note 35.

42 *See. e.g*, Hall, *supra* note 10, at 651 (discussing such a scheme)

43 *See, e.g.,* Marc A. Rodwin, *Patient Data: Property, Privacy & the Public Interest*, 36 AM. J.L. & MED. 586, 593 (2010) (proposing such a scheme).

44 *See* discussion *infra*.

This section briefly clarifies the relationship among data ownership, data commons, and the public domain.

In 2014, the Health Data Exploration Project surveyed a sample consisting primarily of people who track their PHD and found that 54% believe they own their data; 30% believe they share ownership with the sensor company or service provider that enables collection of their data; and 4% believe the service provider owns the data, with only 13% professing indifference.[45] Most respondents viewed ownership as important "because it implies a level of control over the fate of the data," and a significant number of people expressed that they have or would like to have control.[46] Kish and Topol,[47] in their recent call for patient ownership of data, took the stance that individual ownership serves important personal interests that are not served by rights of access and control[48]—a contention this article queries further below.[49]

At the recent Precision Medicine Initiative Summit, President Obama captured many Americans' sentiments: "I would like to think that if somebody does a test on me or my genes, that that's mine."[50] He impressed the lawyers in the room by adding, "But that's not always how we define these issues, right? So there's some legal issues involved."[51] Indeed. Intense feelings of ownership can exist in the absence of legal ownership. Airline passengers feel strong bonds of ownership to their assigned seats and may experience feelings of trespass if a neighboring passenger encroaches on "their" seat. The legal reality is that airline passengers have no property interest in their assigned seats. They merely have a contract for carriage—a type of service contract—with the airline, which can uphold its end of the bargain by reassigning the passenger to any other seat, however unsatisfactory, on the plane or by placing the passenger on the next flight, with or without a voucher toward a future trip on the airline which, by then, one ardently hopes one shall never have to fly on again. We all "like to think" we own our assigned seats and our health data, but this is irrelevant to whether we do.

Discussions of health data commons occur against the backdrop of these calls for patients to own their health data. Terms like scientific research commons, medical information commons, and genomic data commons have roots in a decades-long analysis of commons in natural resource economics and property theory.[52] The scientific and medical communities sometimes refer to "*a* commons" or "*the* commons," apparently to denote a specific, shared data resource, database, or health information infrastructure that would be openly accessible to researchers and clinicians (as if in the public domain), while possibly incorporating respect for individual data ownership. These discussions can seem jumbled to traditional commons scholars, who view

---

45 HEALTH DATA EXPLORATION PROJECT, supra note 35, at 12.

46 *Id*.

47 Kish and Topol, *supra* note 35.

48 *Id*.

49 *See* discussion *infra*.

50 Lily Hay Newman, Obama Says People Who Give Genetic Samples for Research Should Own the Data, Slate (Feb. 26, 2016), at http://www.slate.com/blogs/future_tense/2016/02/26/at_precision_medicine_initiative_summit_obama_says_people_own_their_genetic.html

51 *Id.*

52 *See, e.g.*, ELINOR OSTROM, GOVERNING THE COMMONS (1990) (examining natural resource commons), Carol M. Rose, The Comedy of the Commons: Commerce, Custom, and Inherently Public Property, 53 U. Chi. L. Rev.711-781 (1986) (elaborating the role of commons in property theory more generally).

commons as a multiplicity of possible institutional arrangements[53] (that is, "sets of working rules"[54] for creating, sharing, using, and sustaining a resource) that exist in the space *between*[55] the two extremes of assigning private property rights to the resource or placing it in the public domain.[56] "[T]he knowledge commons is not synonymous with open access" [57] or the public domain, although arrangements that embrace open-access rules are one variety of commons.

Language is a living, evolving thing and members of the scientific and medical communities are free to define the word "commons" in any way that aids their internal communications. It may facilitate broader cross-fertilization of ideas, however, to link discussions of health data commons to other strands of commons analysis. Professors Frischmann, Madison, and Strandburg[58] are modern explicators of the commons analysis identified with Elinor Ostrom and her collaborators in natural resource contexts[59] and later adapted to knowledge resources by Hess and Ostrom.[60] Key points are that commons do not denote specific resources—such as a fishery, pasture, or health database.[61] Commons are not a place or a thing or a resource that people can use.[62] Rather, commons are institutional arrangements for managing or governing the production and use of a particular resource—such as a health database—and for addressing social dilemmas that impede sustainable sharing and stewardship of the resource.[63]

For rivalrous resources, such as a pasture where one group's use may preclude use by others, the number of simultaneously existing commons is often quite limited; some resources can only support the formation of a single commons arrangement. Informational resources are generally conceived as non-rivalrous, because multiple parties can simultaneously use copies of the same information at the same time. Note, though, that *interoperable* health data resources are partially rivalrous, because converting data from disparate sources into well-curated, consistent formats that allow particular types of analysis requires a substantial investment of capital and skilled labor. [64] These latter resources have supply constraints that may limit the number of competing, simultaneous data uses. Data, in theory, could be used for any number of scientific studies, but unless the data are already in an interoperable format (which American data

---

53 See Brett M. Frischmann, Michael J. Madison, and Katherine J. Strandburg, *Governing Knowledge Commons* 1, 3 in GOVERNING KNOWLEDGE COMMONS, *supra* note 3 (illustrating a wide variety of institutional arrangements).

54 Ostrom, *supra* note 52, at 50-51.

55 CHARLOTTE HESS & ELINOR OSTROM, UNDERSTANDING KNOWLEDGE AS A COMMONS (2011).

56 *See* Frischmann *et al.*, *supra* note 53, at 7-8 (contrasting the domain of proprietary rights and the public domain).

57 Hess & Ostrom, *supra* note 55, at 3.

58 Frischmann *et al.*, *supra* note 55, at 3.

59 OSTROM, *supra* note 52.

60 HESS & OSTROM, supra note 55.

61 Frischmann et al., *supra* note 53, at 2.

62 *Id*. at 2.

63 *See id.* at 3 ("Knowledge commons" refers to "the institutionalized community governance of the sharing and, in some cases, creation, of information, science, knowledge, data, and other types of intellectual and cultural resources.")

64 Evans, *supra* note 22, at 102-03.

generally are not)[65], there may not be enough skilled data analysts to convert the data into the formats each study requires.

Generally speaking, though, once data are converted into a common data model or other interoperable format, further uses of the converted data are non-rivalrous. Health data resources thus can support the simultaneous existence of multiple health data commons. For example, a group of three hospitals might form a commons to allow limited data sharing among themselves for use in quality improvement activities. Patients of those same hospitals might merge their own data with data from patients who used other hospitals to form a patient-driven commons to compare their experiences as health care consumers. The two commons, operating under different sets of rules for the benefit of distinctly different groups of users, would co-exist and neither would be *the* commons.

3. The Inevitability of Shared Data Control

In popular culture, property rights are a venerated symbol of an individual's right to restrict other people's access to things that are personally important, such as one's home or one's PHD.[66] Ownership resonates with a new model of privacy that gained ground in recent years. The traditional view of privacy as secrecy or concealment—as a "right to be let alone"[67]—grows strained in an age when the Internet and ubiquitous communication technologies foster broad, voluntary sharing of personal information.[68] We vomit our personal data into the Universe, but we want the Universe to protect our privacy. To conceal the contradiction, modern theorists embrace a new view of privacy in which concealing one's secrets "is less relevant than being in control of the distribution and use by others" [69] of the thick data trails people generate and willingly or unwittingly[70] disseminate. Presently, the "leading paradigm on the Internet and in the real, or off-line world, conceives of privacy as a personal right to control the use of one's data".[71] Property rights, it is hoped, may restore our desired control.[72]

In this debate, the ultimate value of the data ownership metaphor may be its insistence—once property rights are correctly understood—that just and efficient protection of the individual's interests requires limits to individual consent, a proposition embraced less comfortably, when at all, in the bioethics literature. Proposals for individual data ownership

---

65 PRESIDENT'S COUNCIL OF ADVISORS ON SCI. AND TECH., EXEC. OFFICE OF THE PRESIDENT, REPORT TO THE PRESIDENT: REALIZING THE FULL POTENTIAL OF HEALTH INFORMATION TECHNOLOGY TO IMPROVE HEALTHCARE FOR AMERICANS: THE PATH FORWARD 39 (2010) [hereinafter P-CAST REPORT].

66 Sonia M Suter, *Disentangling Privacy From Property: Toward a Deeper Understanding of Genetic Privacy*, 72 GEO. WASH. L. REV. 737, 751 (2004).

67 Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193 (1890).

68 Vera Bergelson, *It's Personal But Is It Mine? Toward Property Rights in Personal Information*, 37 U.C. DAVIS L. REV. 379, 401-402 (2003); Daniel J. Solove, *Conceptualizing Privacy*, 90 CAL. L. REV. 1087, 1092-126 (2002).

69 Raymond T. Nimmer, The Law of Computer Technology, P 16.02, at 16-4, 16-5 (2001).

70 Walter W. Miller, Jr. & Maureen A. O'Roarke, *Bankruptcy Law v. Privacy Rights: Which Holds the Trump Card?*, 38 HOUS. L. REV. 777, 786-87 (2001) (discussing transaction-generated data gathered without consumers' knowledge using "cookies" and related technology).

71 Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815, 820 (2000).

72 Bergelson, *supra* note 68, at 401-02.

sometimes portray property rights as allowing people to veto any unwanted use of their data.[73] This level of control reflects a mythical[74] or fairy-tale[75] view of the legal protections that ownership provides. In practice, property regimes grant individuals a qualified (non-absolute) right to control the disposition of their assets, but they also protect competing individual and social interests by imposing responsibilities and limitations on ownership.[76] Jacqueline Lipton usefully reminds us that ownership actually supplies a bundle of rights, limitations on those rights, and *duties*[77] so that individual and competing interests both receive protection. Margaret Jane Radin sees the two basic functions of property theory as to justify rights and *to explain their boundaries*.[78] As Schlager and Ostrom point out, in natural resource settings, "all rights have correlative duties."[79]

Those seeking absolute control of their PHD should not look to data ownership to give it to them.[80] As an example, homeowners enjoy a very robust set of rights but can be forced without consent to pay property taxes or to cede control of some or all of their property. Their control is subject to easements that may allow public utility projects to cross their land or facilitate a neighbor's access to a landlocked adjacent property. The government has police power to impose duties—such as a requirement for owners to abate hazards or to install sidewalks on their property—that promote public health and safety.[81] Failure to comply may draw sanctions up to and including uncompensated seizure of the property. [82] Eminent domain, or takings, can force non-consenting owners to cede their property for "public use"[83]—a broad concept that in modern law includes commercial office parks and other private endeavors that allegedly confer public benefit,[84] in addition to more traditional public uses such as highways[85]—so long as fair compensation is paid. [86]

---

73 See, e.g., Patricia Mell, *Seeking Shade in a Land of Perpetual Sunlight: Privacy as Property in the Electronic Wilderness*, 11 BERKELEY TECH. L.J. 1, 26-41 (1996).

74 Evans, *supra* note 22, at 77.

75 Suter, *supra* note 66, at 804 (citing Jessica Litman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283, 1297-98 (2000)).

76 Evans, supra note 22, at 77-82; *Jacqueline Lipton, Information Property: Rights and Responsibilities*, 56 FLA. L. REV. 135, 165-75 (2004); Bergelson, *supra* note 68, at 438.

77 Lipton, *supra* note 76 [emphasis added].

78 Margaret Jane Radin, *Property in Personhood*, 34 STAN. L. REV. 957, 958 (1982) [emphasis added].

79 Edella Schlager & Elinor Ostrom, *Property-Rights Regimes and Natural Resources*, 68(3) LAND ECONOMICS, 1992, 249-62, at 250-51 (1992).

80 Evans, *supra* note 22, at 79.

81 John F. Hart, *Land Use Law in the Early Republic and the Original Meaning of the Takings Clause*, 94 N.W.U. L. REV. 1099, 1102, 1107 (2000) (discussing historical uses of the state's police power to require owners to confer positive externalities on the community).

82 Thomas W. Merrill, *The Economics of Public Use*, 72 CORNELL L. REV. 61, 66 (1986).

83 *See* Robin Paul Malloy & James Charles Smith, *Private Property, Community Development, and Eminent Domain* 1, 8, in PRIVATE PROPERTY, COMMUNITY DEVELOPMENT, AND EMINENT DOMAIN (discussing the public use requirement).

84 *See, e.g.,* Poletown Neighborhood Council v. City of Detroit, 410 Mich. 616, 304 N.W.2d 455 (Mich. 1981); Kelo v. City of New London, 545 U.S. 469 (2005).

85 *See* Abraham Bell, *Private Takings*, 76 U. CHI. L. REV. 517, 522 (2009).(noting, "Public ownership of property is not necessary for just and efficient takings.").

86 U.S. Constitution, Amend. V.

The waiver provisions of the HIPAA Privacy Rule and Common Rule strongly resemble the private eminent domain powers that some utility companies enjoy under state laws; any regime of data ownership presumably would incorporate a similar mechanism to ensure access to individuals' data for research that benefits the public.[87] Elsewhere, I have explained why the court-determined fee for a "data taking" would very likely be zero under existing doctrines for assessing takings compensation.[88] Thus, patient data ownership would be unlikely to confer ironclad control or even control superior to what people already have.

Scholars point to additional conceptual and practical flaws with data ownership as a tool to protect privacy: The fact that data resources are largely non-rivalrous undermines the economic justification that property rights are necessary to prevent the waste of scarce resources.[89] Property and privacy serve fundamentally different interests, with property denoting control in the marketplace over things alienable from the self—things we are willing to part with—whereas privacy denotes control over things entwined with our selfhood.[90] Commodifying people's data is potentially objectionable on both moral and practical grounds.[91] Property rights will do little to protect privacy if vulnerable people sell their rights or give them away and, ultimately, things that are owned can nevertheless be stolen. Creating informational property rights is no guarantee that they will function effectively on an ongoing basis; substantial infrastructure and effort may be required to effectuate orderly transfers of "owned" data.[92] The list of objections is long.

Despite these flaws, property terminology is familiar to everyone and the ownership metaphor "will likely stick"[93] in public discourse about data privacy, even if cooler heads resist pressure to enshrine data ownership as law. There is no real harm in invoking property metaphors, as long as we carefully specify what we conceive ownership to mean.[94]

Kish and Topol's recent proposal for patient data ownership, despite its many merits, was not entirely clear what ownership means. They mentioned that "possession is nine-tenths of the law," suggesting a right of access and personal possession, but did not clarify whether this would be a right of *exclusive* possession that allows patients to insist that their healthcare providers erase the provider's copy of their records.[95] Kish and Topol also noted that having a title to one's home creates conditions for trusted exchange, suggesting a right to transfer one's records to others. They listed several conditions to be met: patients should have access to their records anywhere at any time; records should be controlled by the patient or patient's agent; they should be unique and traceable to a real person; records should be privacy-enabled and secure and have a known provenance that allows them to be traced to the data-holder whence they came. These conditions, while useful for purposes of managing data resources, do not all reflect attributes of legal property ownership. Owning one's home does not guarantee that it is secure against break-ins, for example.

---

87 Evans, *supra* note 22, at 79, 85-86.

88 *Id*. at 81.

89 Pamela Samuelson, *Privacy as Intellectual Property*, 52 STAN. L. REV. 1125, 1138-39 (2000).

90 Suter, *supra* note 66, at 737.

91 *Id*. at 799-803.

92 *See* Samuelson, *supra* note 89, at 1136-37 (citing Kenneth C. Laudon, Markets and Privacy, Comm. ACM Sept. 1996, at 92)). *See also* Suter, *supra* note 66, at 804; Litman, supra note 77, at 1297-98.

93 Lipton, *supra* note 76, at 141.

94 *Id*. at 141-42.

95 Kish & Topol, *supra* note 32, at 922.

Some proposals liken data ownership to traditional *fee simple* ownership of a house,[96] but there are many alternative models to consider. Data ownership might, for example, resemble co-ownership in which multiple parties have rights of access and use; it could resemble the nonexclusive rights riparian owners have in a river next to their land—that is, a right to use the river but no right to interfere with others' simultaneous uses such as fishing and navigation;[97] or it could work like a copyright that ends after a period of time and allows fair use by others during that time.[98]

None of the available models of individual data ownership is presently reflected in law. Property law in the United States is set primarily at the state level,[99] except in discrete fields (for example, patent law) where a federal framework controls. Several states have enacted laws that grant individuals a property interest in their own genetic information[100] and a few more states have considered such legislation,[101] although such laws are notoriously vague about what genetic property rights mean.[102] In most states, and for most other categories of health information, state law does not directly address who owns a person's health data. [103]

A few state court cases have found patients own their medical records under specific circumstances.[104] Unfortunately, the pertinent body of state medical records law generally applies in traditional healthcare settings and seemingly does not govern commercial providers of PHD devices and services, such as purveyors of medical and fitness devices. Courts do not recognize an individual property right in personal information such as one's name, address, and social security number.[105] Commercial databases that hold such information are generally treated as the property of the companies that compiled them.[106] In a famous case[107] where plaintiffs

---

96 *See, e.g.,* Mell, *supra* note 73.

97 Eric R. Claeys, *Takings, Regulations, and Natural Property Rights*, 88 CORNELL L. REV. 1549, 1575 (2003). (2003).

98 Abraham Bell, *supra* note 85, at 540-42.

99 Evans, *supra* note 22, at 73.

100 *See* Seth Axelrad, State Statutes Declaring Genetic Information to be Personal Property, *available at:* http://www.aslme.org/dna_04/reports/axelrad4.pdf (listing statutes of Alaska, Colorado, Florida, and Georgia that recognize individual property rights in genetic information). See also, ALA. STAT. § 18.13.010(a)(2); COLO. REV. STAT. ANN. § 10-3-1104.7(1)(a); FLA. STAT. ANN. §760.40(2)(a); GA. CODE ANN. § 33-54-1(1).

101 *See, e.g.,* S.D. H.B. No. 1260 (2012), available at http://legis.sd.gov/docs/legsession/2012/Bills/HB1260P.pdf; Tex. H.B. 2110 (2011), H.B. 1220 (2015), available at http://www.legis.state.tx.us/Search/DocViewer.aspx?ID=84RHB012201B&QueryText=%221220%22&DocType=B.

102 Jennifer K. Wagner & Dan Vorhaus, *On Genetic Rights and the States: A Look at South Dakota and Around the U.S.*, GENOMICS L. RPT., Mar. 20, 2012, available at http://www.genomicslawreport.com/index.php/2012/03/20/on-genetic-rights-and-states-a-look-at-south-dakota-and-around-the-u-s/.

103 David L. Silverman, *Data Security Breaches: The State of Notification Laws*, 19 No. 7 INTELL. PROP. & TECH. L. J. 5, 8 (2007).

104 *See* Susan E. Gindin, *Lost and Found in Cyberspace: Informational Privacy in the Age of the Internet*, 34 SAN DIEGO L. R. 1153, 1195 at n. 231 (1997) (listing several cases where courts have recognized patients' ownership of medical records).

105 David A. DeMarco, *Understanding Consumer Information Privacy in the Realm of Internet Commerce: Personhood and Pragmatism, Pop-Tarts and Six-Packs*, 84 TEX. L. REV. 1013 (2006).

106 *Id*. at 1035-36.

sought to block a company from disclosing their personal information by selling its mailing lists, Vera Bergelson notes an implicit judicial bias "that, to the extent personal information may be viewed as property, that property belongs to the one who collects it."[108] This bias—if it exists—is reminiscent of the ancient *res nullius* doctrine from natural resource law, which treated assets such as subsurface mineral deposits and wild animals as unowned until somebody discovers and captures (takes possession of) them.[109] "Rarely used today, it let private owners stake claims as in the Klondike gold rush."[110]

Law does, however, recognize a subtle difference between a data holder's ownership of a database and its ownership of data that populate the database. "Although it is common for businesses contracting with one another to state that one or another of them 'owns' a particular data set, ownership of the contents of a database is a precarious concept in the United States."[111] Database operators have a legal interest in the data in their databases, but this interest is not usually regarded as ownership.[112] Information in a database can be owned in the sense of being eligible for trade secret protection, if the operator meets all the other legal requirements (such as maintaining the data's secrecy) for such protection.[113] Copyright law typically does not protect database content, which is in the nature of facts rather than expression, so copyright protection typically extends only to features such as the arrangement of a database rather than to the data itself.[114] Even though data holders do not own database content, Marc Rodwin echoes concerns that courts tend to "grant property interests to those who possess that data and preserve the status quo."[115] The status quo is that data are widely scattered in proprietary corporate databases, creating a tragedy of the anticommons that threatens to leave valuable stores of data inaccessible for research and other beneficial uses.[116]

As for PHD generated outside the traditional healthcare setting, the privacy policies and service contracts of device manufacturers do little to clarify data ownership. Scott Peppet recently surveyed privacy policies of twenty popular consumer sensor devices and found only four that discussed data ownership.[117] Three of those four indicated that the device manufacturer, rather than the consumer, owns the data, with some claiming "sole and exclusive" ownership. Such assertions of ownership are not necessarily enforceable at law. Suppose a sensor manufacturer asserted, in its privacy policy, that it owns the Eiffel Tower and a consumer purchased and used the sensor with actual or constructive notice of that policy. Neither of those facts would affect ownership of the Eiffel Tower, if neither party had an enforceable claim to it

---

107 Dwyer v. Am. Express Co., 652 N.E.2d 1351, 1351 (Ill. App. Ct. 1995).

108 Bergelson, *supra* note 68, at 412.

109 John Bouvier, A Law Dictionary, Adapted to the Constitution and Laws of the United States (1856), available at http://legal-dictionary.thefreedictionary.com/Res+nullius.

110 Barbara J. Evans, *Mining the Human Genome after Association for Molecular Pathology v. Myriad Genetics*, 16 GENETICS IN MEDICINE 504, 505 (2014).

111 Silverman, *supra* note 105, at 8.

112 *See* Paul M. Schwartz, *Property, Privacy, and Personal Data,* 117 HARV. L. REV. 2055, 2076-2094 (2004) (discussing conceptual difficulties in applying a regime of property rights to information in databases).

113 *Id*. or Silverman.

114 *Id*.

115 Rodwin, *supra* note 43, at 593.

116 *Id*.

117 Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85, 145 (2014).

under relevant law. Such policies may, however, deter consumer challenges to a manufacturer's alleged ownership.

Even without legal ownership, it is fair to say that health data holders enjoy powers tantamount to ownership. The information they hold is "out of circulation even though it is not, strictly speaking, owned"[118] and many operators "treat patient data as if it were their private property.[119] "Multiple ownership of different pieces of a patient's medical history. . . makes it difficult for anyone to assemble a complete record."[120] The data are siloed.

The reality is that multiple parties have legitimate interests in a person's health data. Healthcare providers must maintain copies of patient data to comply with medical records laws, to ensure continuity of care during patients' future visits, to defend against lawsuits. Insurers need records for auditing, fraud prevention, and state regulatory purposes. Even in theory, exclusive data ownership is unworkable. If data ownership existed, it seemingly would have to be some form of shared ownership.

What might shared data ownership look like? In a different context, Ostrom and Schlager identified a set of entitlements individuals enjoy in shared property regimes for natural resources, such as fisheries.[121] These resonate with some of the entitlements people wish to have in relation to their PHD. The list includes "operational-level" entitlements:

(1) a right of access to the resource, and
(2) a right to withdraw products from the resource (e.g., a right to catch fish), corresponding to a right to use data resources.[122]

A party may—but does not always—have additional "collective-choice" rights, including:

(3) a right of management, which confers the right to participate in decisions about resource uses and the right to improve or transform the resource (such as by adding, deleting, or editing data),
(4) a right of exclusion, which confers the right to participate in decisions about who can access and use the resource and decisions about the appropriate process for approving and enforcing access and use, and
(5) a right of alienation, to transfer the above rights to other people.

In shared ownership regimes, the individual does not usually enjoy "sole and despotic dominion,[123] such as an unassailable right to consent to, or veto, specific resource uses. Rather, the individual has a voice (voting rights) in a collective decision-making process, and a residual right to exit the collective if its decisions prove unsatisfactory.

4. The Challenge of Assembling Large-Scale Data Resources

In traditional healthcare and research environments, control of data remains fragmented among multiple data holders (such as physicians, research institutions, hospitals, insurers),[124] with another layer of fragmentation at the level of individuals, to the extent their consent is required

---

118 Hall, *supra* note 10, at 646.
119 Rodwin, *supra* note 43, at 588.
120 Hall, *supra* note 10, at 647.
121 Schlager & Ostrom, *supra* note 79, at 250-51.
122 *Id.* at 250.
123 2 WILLIAM BLACKSTONE, COMMENTARIES *2, *available at* http://avalon.law.yale.edu/18th_century/blackstone_bk2ch1.asp (spelling conformed to modern conventions).
124 Hall, *supra* note 10, at 647-48; Rodwin, *supra* note 43, at 606.

for data access.[125] As tracking and surveillance technologies and direct-to-consumer testing services generate PHD outside traditional healthcare settings, the fragmentation increases, with different data holders operating subject to different legal and regulatory regimes.[126] Harnessing data for public good requires transactions to bring the data together. Subpart A of this discussion discusses traditional healthcare and research environments. Subpart B then highlights key differences affecting non-traditional PHD such as self-tracking sensor data.

A. Traditional healthcare and research data

The Common Rule and HIPAA Privacy Rule frequently apply in traditional clinical care and research settings. These regulations endow individuals and data holders with various entitlements,[127]somewhat resembling the entitlements Schlager and Ostrom associate with shared resource ownership regimes.[128] Note that these are regulatory entitlements, not ownership rights, but they include various rights of access, use, management, exclusion, and alienation and, in some respects, they are "strikingly similar" to a scheme of data ownership.[129] By exercising their regulatory entitlements, the parties may be able to free data for use in assembling large-scale data resources for research and public health.

When a data holder stores a person's data, both parties share control and their interests are not necessarily aligned. Individual consent may not be sufficient to ensure data access. In this respect, assembling data for informational studies is fundamentally different from the problem of enlisting human participants for clinical (interventional) studies. Access to the key resource for interventional research (human beings to study) is properly modeled as a two-party transaction between an individual and a prospective user such as an investigator who desires to involve the individual in the research. The individual has exclusive control over the resource—the individual's body—to which the prospective user needs access. Assuming the individual has decisional capacity and is willing to consent, the consent is sufficient to ensure access. In contrast, acquiring the key resource for informational research (data) is poorly modeled as a two-party transaction. It often requires three-way transactions among the prospective data user, the data subject, and a data holder that possesses the person's data. Figure 1 displays four possible solutions to the problem of data access.

Solution 1: Incentivized consent alignment. In Quadrant 1, data access is unproblematic because there is consent alignment. The individual and the data holder both want to share the individual's data. Consent alignment sometimes arises naturally, but it also may be possible for policy makers to create incentives for the parties to align.

Incentivized consent alignment has a proven track record in research settings. Conditional grants are an effective mechanism. For example, the National Institutes of Health (NIH) and counterpart funding agencies in other nations have encouraged sharing of genomic data by implementing policies that require grantee research institutions to deposit certain data they generate under grants into shared genomic databases.[130] If the deposited data are de-identified,

---

125 Rodwin, *supra* note 43, at 606.

126 *See* discussion *infra* at subsection B.

127 Evans, *supra* note 22, at 82-86..

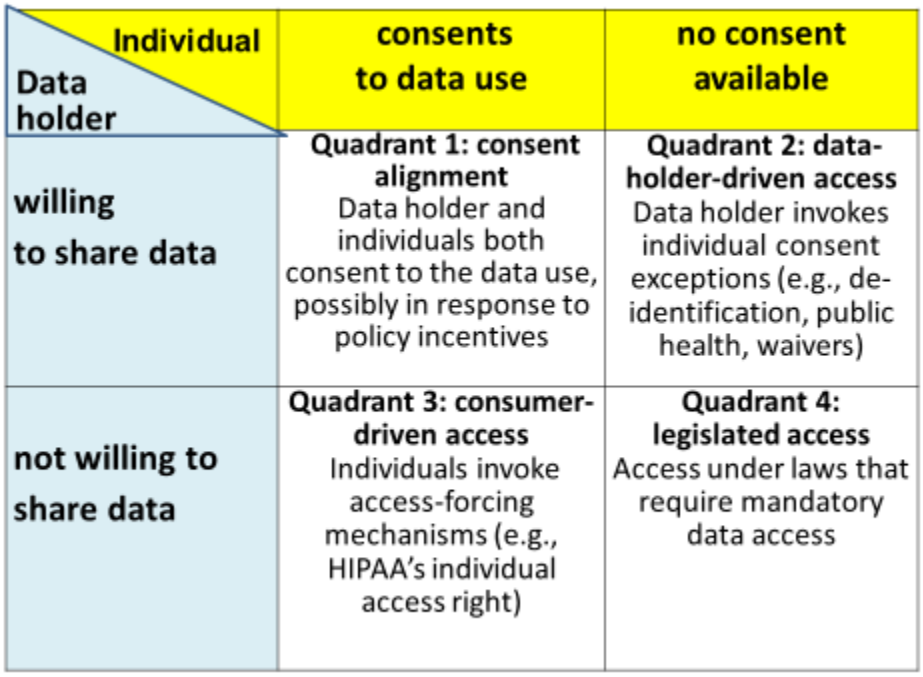
128 *See* discussion *supra*.

129 Evans, *supra* note 22, at 82.

130 *See generally* Evans, Burke & Jarvik, *supra* note 19 (discussing NIH data-deposit policies) and Jorge L. Contreras, *Constructing the Genome Commons* 99-119, in GOVERNING KNOWLEDGE COMMONS, *supra* note 3 (discussing funding agencies' data-sharing policies)

individual consent is not required by current regulations. If a research project contemplates sharing data in a form that does need individual consent (for example, sharing of data among

Figure 1: Pathways for Assembling Large-Scale Data Resources



participating institutions in a multi-site study), consent can be procured at the point when participants consent to the research.

The NIH-funded Precision Medicine Initiative's million-person cohort is another example incentivized consent alignment. Data holders and the individual data partners who wish to participate in this exciting project will be asked to consent to its data-sharing terms.[131] Publicly funded efforts of this sort can jump-start discovery and pave the way for future efforts. The major drawback is that conditional funding solutions are costly and, amid budgetary constraints, raise concerns about long-term sustainability. They are not scalable as a way to develop the data resources ultimately required for 21st-century science, which may need to include hundreds of millions of people.

Consent alignment is more difficult to achieve in clinical and commercial healthcare settings, yet it is potentially critical to the overall effort. These environments hold large stores of clinical health data that are essential for assembling deeply descriptive data resources that link individuals' phenotypes to their genotypes. Clinical laboratories and healthcare providers have various commercial incentives not to share data that they hold.[132] The data-deposit policies of

---

131 DJ Patil, Claudia Williams, Stephanie Devaney, Your data in your hands: Enabling access to health information, (Mar. 10, 2016), at https://medium.com/@WhiteHouse/your-data-in-your-hands-enabling-access-to-health-information-6fce6da976cb#.60242uv2i.

132 R. Cook-Deegan, J.M. Conley, J.P. Evans and D.Vorhaus, *The Next Controversy in Genetic Testing: Clinical Data as Trade Secrets?* 21 EUROPEAN JOURNAL OF HUMAN GENETICS 585-588, at 585 (2013).

funding agencies like NIH are not binding on these data holders, although many do voluntarily contribute at least some data to shared public data resources like ClinGen/ClinVar.[133] Commercial data holders' reluctance to share data poses an important barrier to the assembly of large-scale, linked data resources, even as surveys show most individuals would be willing to share.[134]

Policy options for fostering consent alignment—including in clinical and commercial settings—have not been exhausted. Policy-makers should explore approaches for incentivizing consent alignment. Possible approaches include conditioning desirable benefits (as opposed to funding) on data sharing. Commercial data holders may voluntarily agree to share data if a benefit, such as Medicare reimbursement or Food and Drug Administration (FDA) approval of a medical product, depends on it. Such approaches may fail to achieve alignment, however, if individual consent is also required. Incentivizing individual consent, depending on the context, may or may not raise concerns about coercion, so incentives must be thoughtfully designed.

Solution 2: Data-holder-driven access. Consent alignment can fail in two ways, portrayed in Quadrants 2 and 3 of Figure 1. In Quadrant 2, the individual is unwilling to share data or cannot practicably be reached to provide consent and privacy authorization. The data holder is willing to share. Problems of this sort seem destined to occur in the future, if growing awareness of re-identification risks makes individuals wary of consenting to research with their data.

The HIPAA Privacy Rule and Common Rule pragmatically address the situation in Quadrant 2 with an array of individual consent exceptions, exemptions, and definitional loopholes that protect society's interest in enabling certain important data uses. These exceptions permit, but do not require, a data holder to release data and, for that reason, Figure 1 characterizes these pathways as "data-holder-driven" access mechanisms. Data can be freed through these legal pathways, but only if the data holder wants to do so.

For example, the Privacy Rule and Common Rule allow data holders to supply data, without individual consent, for certain public health, regulatory, and judicial uses of data,[135] but the regulations do not themselves require data holders to share. In one recent study, IRBs refused to provide about 5% of the requested medical records for a well-documented, congressionally authorized public health purpose.[136] Users that are denied data access must look to other sources of law—such as a court order or provisions of state public health laws that *require* healthcare providers to report specific types of information—to force data access. HIPAA and the Common Rule do not themselves mandate access for these uses.

This is also true of their waiver provisions,[137] which let Institutional Review Boards and privacy boards (collectively, "IRBs")[138] approve unconsented research uses of data—including,

---

133 Evans, Burke & Jarvik, *supra* note 18.

134 *See supra* note 35.

135 *See, e.g.*, 45 C.F.R. section 164.512 (listing HIPAA exceptions).

136 SARAH L. CUTRONA, ET AL., MINI-SENTINEL SYSTEMATIC VALIDATION OF HEALTH OUTCOME OF INTEREST: ACUTE MYOCARDIAL INFARCTION CASE REPORT, 10, 12 (2010), *available at* http://www.mini-sentinel.org/work_products/Validation_HealthOutcomes/Mini-Sentinel-Validation-of-AMI-Cases.pdf.

137 45 C.F.R. § 164.512(i) (2010) (HIPAA waiver provision); *id.* § 46.116(d) (Common Rule waiver provision).

138 *See* 45 C.F.R. §§ 46.103(b), 46.107–108 (describing IRBs: private ethical review panels, often staffed by employees of the data holder or data-using research institution, to which the Common Rule delegates various aspects of research oversight); *id.*§ 164.512(i)(2)(iv) (allowing of waivers of

crucially, data with identifiers that can be linked across separate datasets to form longitudinal health records. [139] When IRBs are affiliated with the data holding institution, this effectively allows the data holder to override individual consent. When an external IRB approves a waiver, the data holder may, but is not required to, release the data,[140] so discretion still rests with the data holder. Unfortunately, as already noted, data holders do not always wish to share.

Data-holder-driven data access under HIPAA and the Common Rule has been an important pathway for enabling data for important research and public health uses. It will not suffice, however, as a way to assemble the large-scale, deeply descriptive data resources that the future requires. IRBs may be comfortable concluding that conditions for waiving consent (such as that privacy risks are minimized) are met in the context of a discrete proposed data use. Certifying that these conditions are met is far more difficult, however, in the context of a large-scale data infrastructure that will be widely accessible for many different uses. Moreover, even in past contexts where data-holder-driven access has worked, it has never been uncontroversial. Bioethicists and data subjects criticize its disrespect for the individual's right of choice.[141] Law scholars criticize the democratic illegitimacy, procedural deficiencies, and potential conflicts inherent in using IRBs as the decision-makers to override individual consent.[142] Investigators and institutions find it cumbersome to administer.[143] There is little to like about this method of access, particularly in contexts such as Precision Medicine that aim to empower patients and research subjects. Something new is needed.

Solution 3: Consumer-driven data access. Consumer-driven access may be the needed alternative. There have been isolated instances, such as the Canavan example discussed in the introduction, where patient advocacy groups took the lead in assembling data resources for research into specific diseases. The question is whether such efforts are scalable: Could they be used to assemble larger data resources for more general use in diverse research contexts, under terms and conditions set via consumer self-governance?

In Quadrant 3 of Figure 1, the individual wishes to share his or her stored data, but this desire is thwarted by an unwilling data holder. The bioethical literature is asymmetrical, evincing concern about unconsented uses of people's data (Quadrant 2), while largely failing to register ethical objections when data holders or their IRBs block data access for uses of which the individual would have approved (Quadrant 3). The Common Rule exemplifies this asymmetry: It

---

consent under the HIPAA Privacy Rule to be approved by either a Common Rule-compliant IRB or by a HIPAA-compliant "privacy panel" that is similar to an IRB).

139 *See, e.g.,45 C.F.R.* § 164.512(i) (providing, in the HIPAA waiver provision, that uses and disclosures pursuant to a waiver are "permitted" — i.e., disclosures are allowed but not required); *id.* § 46.116(d) (couching the Common Rule's waiver provision in similarly permissive language: "An IRB may approve . . . ."). The IRB of a research institution that wishes to receive data from a data holder can approve a waiver authorizing release of the data. *See* Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,695 (Dec. 28, 2000) (rejecting, in the preamble to the HIPAA Privacy Rule, suggestions that HHS should require IRBs that approve waivers to be independent of the entity conducting the research). Under section 164.512(i), recipient-approved waivers permit the data holder to disclose data but do not require it, so the recipient has no way to force the provision of needed data and services.

140 *Id.*

141 IOM, Privacy Report, *supra* note 29.

142 Coleman, *supra* note 40.

143 IOM, Privacy Report, *supra* note 29.

contains multiple consent exceptions[144] allowing data holders to share people's data without their consent in Quadrant 2, but contains no access-forcing mechanism to help data subjects free their data for socially valuable uses in Quadrant 3. This defect may be unintentional, an artifact of cramming three-party data access transactions into the Common Rule's simplistic two-party model of human-subject enrollment in clinical trials.

The HIPAA Privacy Rule, in contrast, contains an access-forcing mechanism. The Privacy Rule was expressly designed for data transactions (and, in particular, those where a person's data are held by an institutional or corporate data holder). It takes a more symmetrical approach that facilitates data access in Quadrant 3 as well as in Quadrant 2. Section 164.524 of the Privacy Rule grants individuals a right of access to information about themselves that a HIPAA-covered entity holds in its files.[145] By exercising this right, individuals can obtain their data[146] and then, if they wish, contribute it for research and other uses. Mark Hall once proposed a scheme in which a patient could grant a license to a trusted intermediary, which would exercise the patient's Section 164.524 access rights to gather the patient's data from the various HIPAA-covered healthcare organizations that hold portions of the patient's medical data, assemble the data into a longitudinal record, and then act as the patient's agent for purposes of negotiating access with researchers and other prospective users in accordance with the patient's preferences.[147] This vision is now set to become reality.

In 2014, the U.S. Department of Health and Human Services amended[148] the Privacy Rule and the Clinical Laboratory Improvement Amendments of 1988 (CLIA) regulations[149] to expand the reach of people's section 164.524 access right to include information held at HIPAA-covered clinical laboratories.[150] Laboratories had not previously been subject to the access right, which applied to other healthcare providers such as hospitals and clinics. The changes became legally effective on October 6, 2014 amid considerable confusion and even disbelief about the apparent scope of the new individual access right. A legal analysis by an NIH-funded working group concluded that at HIPAA-covered laboratories that conduct genomic testing, whether for research or clinical purposes, the accessible dataset includes not just final, interpreted test reports but also underlying genomic data if the laboratory has stored data in a form that is traceable to the requesting consumer. Patients who sought to exercise their new rights of access to laboratory-held information after October 2014 sometimes encountered barriers, however, and were not able to access their data.

The HHS Office for Civil Rights, which administers the Privacy Rule, issued guidance in January and February of 2016 confirming that the Section 164.524 access right applies to underlying genetic variant data generated during genomic sequencing, as well as to finished test reports which typically focus on just a few target variants; that individuals have a right to request their data in machine-readable (electronic) format; and that they can direct the data holder to

---

144 *See supra* notes 34-36, 137-39 and accompanying text.

145 45 C.F.R. § 164.524. See also Barbara J. Evans, Michael O. Dorschner, Wylie Burke & Gail P. Jarvik, *Regulatory Changes Raise Troubling Questions for Genomic Testing*, 16 GENETICS IN MEDICINE 799-803 (2014). (discussing the scope of this access right at genomic testing laboratories).

146 Evans, Dorschner, *et al.*, *supra* note 145.

147 Hall, *supra* note 10, at 650, 660-61.

148 U.S. Dep't of Health and Human Servs., CLIA program and HIPAA privacy rule; patients' access to test reports, 79 Fed. Reg. 7290 – 7316 (Feb. 6, 2014).

149 42 C.F.R. pt. 493

150 79 Fed. Reg. at 7292.

transfer data to an agent or trusted intermediary of their choosing. Consumers now have the power to force HIPAA-covered laboratories to release data from clinical testing. Moreover, they also can access data generated at HIPAA-covered research laboratories (such as laboratories embedded in HIPAA-covered academic medical center).[151]

The section 164.524 access right was originally conceived as an instrument to enhance privacy protection. In the preamble to the original Privacy Rule in 2000, HHS cited a "well-established principle" that an individual (or designated personal representative) should have "access rights to the data and information in his or her health record"[152] and remarked that people's "confidence in the protection of their information requires that they have the means to know what is contained in their records."[153] More recently, HHS acknowledge that the individual access right has broader importance. In its 2014 rulemaking, HHS described section 164.524 as crucial not merely to enhance privacy protection but also because: (1) it "enable[s] patients to have a more active role in their personal health care decisions"; (2) it is consistent with "certain health reform concepts" including personalized medicine, participatory medicine, disease management and prevention; and (3) it supports HHS's goals and commitments regarding widespread adoption of electronic health records (EHRs).[154] These last two points conceive section 164.524 as an instrument to free data for public good.

The Privacy Rule pits the rights of the individual and the rights of the data holder against one another, and the interplay/tension between the two helps protect the public's interest in data access. This is far from a perfect access scheme—for example, the data resources available in Quadrant 3 may be marred by selection bias, because these datasets only include individuals who took volitional steps to free their data for research.[155] Access is nevertheless broader than under the lopsided framework of the Common Rule.

Solution 4: Mandatory data sharing. This leaves Quadrant 4, the all-too-common situation where neither the data subject nor the data holder is motivated to share data for the public good. Data users have little recourse when this is true. As noted above, Common Rule and HIPAA Privacy Rule contain no provisions *requiring* data holders to free data, other than HIPAA's section 164.524 access right. Authority to force access must come from other sources of law. Legislatures offer a legitimate, democratic mechanism for imposing binding, collective decisions on data holders and individual data subjects who, in Quadrant 4, dissent from a socially beneficial data use. Access-forcing laws are, however, very rare and typically focus on narrow problems where the need for access is compelling (e.g., reporting of child abuse and communicable diseases). They generally do not address the problem of freeing data for research.

---

[151] U.S. Dep't of Health & Human Servs., Individuals' Right under HIPAA to Access their Health Information 45 CFR § 164.524, at http://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/ and Questions and Answers About HIPAA's Access Right, at http://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/#newlyreleasedfaqs.*See also* Evans, Dorschner *et al*, *supra* note 147.

[152] U.S. Dep't of Health & Human Servs., Off. of Sec'y, Off. of Ass't Sec'y for Planning & Eval., Standards for Privacy of Individually Identifiable Health Information: Final Rule, 65 Fed. Reg.82462, 82606 (Dec 28,2000) (to be codified at 45 C.F.R. pts. 160, 164) (Dec. 28, 2000).

[153] *Id*. (quoting National Committee for Quality Assurance and the Joint Commission on Accreditation of Healthcare Organizations).

[154] 79 Fed. Reg. at 7290.

[155] *See supra* note 29 and accompanying text (discussing selection bias).

Forcing access to data in Quadrant 4 presents difficult legal issues, which may explain lawmakers' reluctance to require mandatory access. Forcing private-sector data holders to disclose their data may constitute a taking and require "just compensation,"[156] if courts recognize the data holders' asserted ownership of the data (or of the capital they invested to marshal the data and develop their health data infrastructures). A related problem is that creating useful research data resources requires inputs not just of data, but of services (such as to convert data to an interoperable format) that data holders would need to provide).[157] The government has little power to force data holders to contribute services,[158] even if they could be forced to share their data. The needed services can only be procured by consensual methods, such as entering contracts with the data holders[159] or requiring their services as a condition of a grant,[160] all of which require funding. Legislatures, when enacting access-forcing other laws, would need to provide the necessary funding. This, again, explains why access-forcing legislation is rare. Forced access is a possible solution in specific, narrow contexts, but it is not a solution to the broad problem of assembling the large-scale, deeply descriptive data resources for 21st-century science.

B. Special challenges with non-traditional PHD

Even within the category of wearables and other sensors, "the information from two disconnected sensing devices can, when combined, create greater information than that of either device in isolation."[161] An example of this "sensor fusion" is that data on heart rate and respiration, when combined, may support inferences about substance abuse.[162] Linking sensor data that reflect lifestyle, exposures, and environment to traditional health and genomic data would be even more powerful, and that is a goal of developing 21st-century data resources.

Harnessing non-traditional PHD for public good requires a framework to support multi-party consent transactions, as modeled in Figure 1. Individuals can exercise unilateral control over PHD stored on their own sensor devices, but much of their PHD may be externally stored and subject to full or partial control of a device or sensor manufacturer, service provider, or other data holder ("PHD company"). PHD companies rarely are bound by the HIPAA Privacy Rule and Common Rule. Except where isolated state-law privacy protections apply to them, their privacy and access policies are largely a matter of company policy. Whether data are stored on the device, by the company, or by both depends, of course, on the device, how much storage capacity it has, and on the service contract that accompanies the device. Many companies do act

---

156 U.S. Constitution, Amend. V.

157 Evans, *supra* note 22, at 102-03.

158 *See* Susan W. Brenner with Leo L. Clarke, *Civilians in Cyberwarfare: Conscripts*, 43 VAND. J. TRANSNAT'L L. 1011, 1056–57 (2010) (discussing whether the government can require civilian workers to perform services aimed at protecting against cyberterrorism and finding only one example, during the Revolutionary War, where Congress forced civilians to provide services outside the context of conscripted military service).

159 *See* Steven J. Kelman, *Contracting*, *in* THE TOOLS OF GOVERNMENT 282, 283–85 (Lester M. Salamon ed., 2002) (discussing features of contracts through which the government procures products or services for its use); *see also* Ruth Hoogland DeHoog & Lester M. Salamon, *Purchase-of-Service Contracting*, *in* THE TOOLS OF GOVERNMENT *supra*, 316, 320 (describing contracts in which the government procures services for delivery to third parties such as beneficiaries of welfare programs).

160 *See* David R. Beam & Timothy J. Conlan, *Grants*, *in* THE TOOLS OF GOVERNMENT, *supra* note 159, at 340, 341 (discussing the government's use of grants to stimulate performance of services).

161 Peppet, *supra* note 117, at 93.

162 *Id*.

as data holders for their customers and, for many devices, the company is the primary holder of the consumer's data. There are competing interests: the consumer, the data holder, and the public as represented by researchers and other users who desire access to the data for socially beneficial purposes. Balancing these interests requires a scheme of informed consent, well-tailored consent exceptions, and one or more access-forcing mechanisms.

Many observers think it unlikely that legislators will act to create a privacy and access scheme for PHD companies. Congress has broad power to set rules under the commerce clause and presumably could establish rules for PHD companies if it wanted to. Yet Congress has refused for many decades to impose a uniform ethical and privacy framework on private-sector (non-publicly-funded) research activities. The Common Rule is mandatory only for research that the federal government funds (or at institutions that receive public funds, if proposed changes go into effect. The Common Rule is an exercise of Congress's spending power, which regulates by placing strings on gifts of federal funds. PHD companies, firmly rooted in the private sector, is not in the habit of taking federal funds and therefore lacks the "hook" that Congress has used to regulate privacy and ethical issues in traditional academic research settings.

The Federal Trade Commission (FTC) and state attorneys general have jurisdiction to regulate unfair or deceptive business practices and can act, for example, if a PHD company publishes a privacy policy that it later dishonors.[163] The recent Wyndham appeal[164] recognized that FCC has authority to regulate cybersecurity more broadly even if a company has not dishonored a policy it previously published. The FTC is actively engaged in efforts to protect consumer information privacy and has published reports and Fair Information Practices to guide Internet and consumer-data companies.[165] Congress has not, however, made them mandatory.[166]

The cautious assumption is that the privacy of, and access to, PHD will continue to be governed largely by company policies and voluntary industry self-regulation. A survey of such policies by Scott Peppet found that some data-holding PHD companies promise to not to share consumers' personally identifying information (PII).[167] As with traditional health data, re-identification is a growing concern with sensor data.[168] Prof. Peppet cites an intelligence source for the proposition that if fitness data reveals the gait at which a person walks, unique identification may be possible.[169] PHD companies' policies tend to be vague when defining PII: does it merely include consumers' names and other overt identifiers or is re-identifiable sensor data also PII?[170] Companies generally reserve the right to share or sell non-personal information (non-PII) more broadly than PII, but their policies may leave it unclear which data are accessible for research.[171] A reasonable expectation for consumers, absent a clear policy to the contrary, is that a company's promise not to share PII amounts to a promise to strip consumers' sensor data

---

163 DeMarco, *supra* note 105, at 1040-41.

164 Federal Trade Comm'n v. Wyndham Worldwide Corporation, 799 F.3d 236 (US Court of Appeals, Third Circuit, August 24, 2015).

165 *See, e.g.,* U.S. Federal Trade Comm'n, Protecting Privacy in an Era of Rapid Change: Recommendation for Businesses and Policymakers (March 2012); U.S. Federal Trade Comm'n, Internet of Things: Privacy and Security in a Connected World (January 2015).

166 DeMarco, *supra* note 105, at 1040-41.

167 Peppet, *supra* note 117, at 129.

168 *Id.*

169 *Id*.

170 *Id*. at 143-44.

171 *Id*. at 144.

of overt identifiers before it is shared. Despite the theoretical risk of re-identification, removing overt identifiers affords at least some privacy protection, and consumers still display considerable willingness to have their data shared for research in de-identified form.[172]

Professor Peppet's survey of found PHD companies' policies were "likewise inconsistent in the access, modification, and easy mechanism for exportation of raw sensor data."[173] In contrast to the HIPAA-regulated space, PHD consumers have limited access rights, and their access rights are often indefinite. Some companies' policies allow access to PII but not to non-PII. When these terms are vaguely defined—as often happens—the scope of a consumer's access right may be quite limited. Debjanee Barua et al. found that consumers want to be able to get a copy of their data: "This is the simplest level of control over one's data—the ability to inspect, manipulate, and store your own information. But it's not usually possible."[174]

Legislation or regulations may not be the best or the swiftest way to fix these problems. The PHD industry is responsive to its consumers and the consumers tend to be an educated, empowered group.[175] The best path forward may be to educate consumers about appropriate privacy, data security, and access standards; to mobilize consumers to demand such standards; and develop a system of voluntary certification for PHD companies that makes it easy for consumers to identify those that implement the standards consumers demand. The question of what the standards should be is beyond the scope of this article and, in any event, should be resolved with meaningful input from consumers themselves.

One point, however, is very clear: Access and transfer rights, as broad and enforceable as the Privacy Rule's section 164.524 provides, are a crucial access-forcing mechanism to free data for transfer into consumer-driven data commons. There are costs of providing an access mechanism and many practical issues, such as making sure consumers can access their data in useful formats.[176] As with HIPAA's access right, PHD companies should be able to charge a reasonable, cost-based fee for servicing access requests, but such fees should be subject to rules concerning what the fees can include.[177] For FDA-regulated devices, access to one's own data potentially bears on a device's safety and effectiveness. FDA should explore whether it has authority to impose minimal consumer access rights on devices the agency regulates.

5. The Necessity of Collective Self-Governance

Discussions of governance of large-scale data commons all too often conflate governance models with system architecture. That is a distraction. The crucial question in governance is not whether to establish a central database versus a federated/distributed data network or a network of networks. Rather, governance is about control relationships: Who gets to decide whose data will be included in a large-scale data resource, the rules of access to the data, the list of permissible uses and terms of use, the privacy and security protections, and the procedures for making such decisions? A consumer-driven data commons is one in which such decisions would be made collectively by the people whose data are involved.

---

[172] *See supra* note 37.

[173] Peppet, *supra* note 117, at 145-146, n. 355.

[174] *Id*. at 161-162.

[175] HEALTH DATA EXPLORATION PROJECT, *supra* note 35.

[176] Peppet, *supra* note 117 (citing DeMarco, *supra* note 105, at .1042).

[177] *See* U.S. Dep't of Health & Human Servs., Questions and Answers About HIPAA's Access Right, at http://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/#newlyreleasedfaqs (discussing the reasonable cost-based fee).

Consumer-driven data commons would grow up alongside—and, if they succeed, possibly replace over time—the data-holder-centric models of the past, which grow increasingly unworkable in the environment of big data, where de-identification is dead. Consumer-driven data commons also will exist alongside consumer-driven access models that rely on granular individual consent to specific data uses. In granular consent models, individuals exercise their access rights to free data from data holders, but there is no collective governance of the overall data resource and each individual makes decisions strictly for herself. Granular consent has many merits—such as overcoming people's initial reluctance to contribute their data and fostering a partnership—and it is the model the Precision Medicine apparently is embracing for its initial million-person research cohort.

Long-term, however, a granular individual consent has a fatal flaw: The grand scientific challenges of the 21st century require collective action to resolve, and granular individual consent fractures the people, limits us and makes us small by robbing us of the capacity for collective action. It consigns us to the state Thomas Hobbes referred to as "the confusion of a disunited multitude,"[178] unable to act together to conquer grand challenges. In Hobbes' scheme, the greater power of a Commonwealth is instituted when a multitude of people come together and covenant "every one with every one," to create institutions for making collective decisions, so that "every one, as well he that voted for it as he that voted against it" shall embrace decisions made by the "consent of the people assembled … in the same manner as if they were his own."[179] Consumer-driven data commons are vehicles for groups of consenting individuals to work together to build more inclusive datasets than their members, acting alone, could offer for scientific use. Members might further enhance the inclusiveness of their data assets through transactions to merge their own data resources with those of other commons-forming groups.

Whatever the possible merits of universal, compulsory contribution of individual PHD, that is not what consumer-driven data commons are about. Rather, they are smaller, self-governing groups of *consenting* individuals, who have rights to enter and leave the commons on transparent terms that each commons-forming group would itself establish. People would still have a right to consent, but it would be a right to enter, or not enter, a specific commons arrangement. Those choosing to place their data in a consumer-driven commons would, thereafter, have collective choice rights to participate in decisions about how the entire data resource—*including the data of all members*—can be used. Members could elect to leave the commons, but while members, they would be bound by its collective decisions regarding permissible uses of their data.

As consumer-driven commons groups develop their own rules of access and use, privacy practices, decision-making processes, they offer a laboratory for modernizing ethical norms to accommodate the age of big data. The existing bioethical norms surrounding informational research were heavily influenced by norms designed for the clinical or interventional research setting. The strong norm of individual, protocol-specific informed consent in interventional research has deep roots in the common law notion that unconsented touching of a person's body constitutes battery. Unauthorized invasions of the human body are "offensive to human dignity"[180] and our legal system has long credited the "the right of every individual to the possession and control of his own person, free from all restraint or interference of others."[181] But

---

[178] THOMAS HOBBES, LEVIATHAN 101.

[179] *Id*. at 101.

[180] Rochin v. California, 342 U.S. 165, 174 (1952).

[181] Union Pac. Ry. Co. v. Botsford, 141 U.S. 250, 251 (1891).

does it follow that touching a person's data is equivalent to touching the person's body, so that the same consent norms should apply in informational research?

The clamor for individual data ownership—which is widely misperceived as conferring inviolable individual consent rights--often draws inspiration from John Locke or from the concept of property as an aspect of personhood.[182] Yet the Lockean concept that people own their bodies does not imply that they own data about their bodies, just as home ownership does not imply ownership of house-related information, which is widely available to realtors, taxing authorities, building inspectors, and busybodies curious about its square footage, improvements, and market value. People whose workouts generate fitness tracking data undoubtedly have earned "sweat equity" in their PHD under Locke's labor theory of ownership, but the company that designed and marketed the device and invested effort to capture and store the PHD could assert an equally plausible claim under this theory.[183]

The personhood theory of ownership recognizes a moral claim to things that are integrally related to one's self-development and sense of personhood.[184] Much—some say too much—has been made of the "tenuous link between personal information and personhood."[185] The "Quantified Self" name of a collaboration of users and makers of self-tracking tools[186] and the "Welcome to You" greeting on 23andMe's web site[187] are marketing claims, not ontological claims. Your data are not actually who you are.

Philosopher Charles Taylor bemoans the fact that modern discourse has banished ontological accounts of human worth from the discussion of morality.[188] "Ontological accounts have the status of articulations of our moral instincts. … If you want to discriminate more finely what it is about human beings that makes them worthy of respect, you have to call to mind what it is to feel the claim of human suffering, or what is repugnant about injustice, or the awe you feel at the fact of human life."[189] The claim that PHD is integral to selfhood lends credence to Taylor's alarm about impoverished ontological accounts of the modern Self. Early assertions that genetic information is integral to selfhood[190] credited genomic science with a predictive power that, it is now clear, was fanciful. A person's genome is a "future diary" recorded in a largely foreign language with most of the words inscrutable.

Research and public health uses of data are not directed at reading personal secrets in people's diaries. As Lawrence Lessig points out, "No one spends money collecting these data to actually learn anything about you. They want to learn about people like you."[191] The variants in one's genome are interesting to scientists only insofar as the variants are shared by other people and, if they are shared, in what sense can one person "own" them? Genomic testing has been

---

182 Radin, *supra* note 78, at 958, n. 3.

183 See DeMarco, *supra* note 105, at 1036, n 109 (arguing that Locke's labor-desert theory supports the notion that those who invest to collect data have rights in it).

184 *Id.*

185 DeMarco, *supra* note 105, at 1014.

186 http://quantifiedself.com/about/

187 *See* 23and Me Homepage, available at https://www.23andme.com.

188 CHARLES TAYLOR, SOURCES OF THE SELF: THE MAKING OF THE MODERN IDENTITY 5 (Harvard University Press 1992).

189 *Id*. at 8 (emphasis added).

190 *See, e.g.,* George J. Annas, *Privacy Rules for DNA Databanks: Protecting Coded 'Future Diaries,'* 270 JAMA 2346 (1993).

191 LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 152 (1999).

misrepresented as "intensely private" when, in fact, the genome is a public space—perhaps the ultimate public space. It is where we go to discover what we have in common with other people.

Ruth Faden et al. acknowledge that the "moral framework for a learning healthcare system will depart in significant respects from contemporary conceptions of clinical and research ethics"[192] and may include an obligation for patients to participate in learning activities.[193] Faden and her coauthors see this as a bounded obligation that would vary with the level of burden and risk involved. While not obligated in risky clinical trials, people may have an ethical duty to contribute their data to studies that can advance useful knowledge while providing reasonable data security.[194] Faden et al. suggest that this obligation is justified by a "norm of common purpose… a principle presiding over matters that affect the interests of everyone."[195] "Securing these common interests is a shared social purpose that we cannot as individuals achieve."[196] The notions of common purpose, security common interests, and shared social purpose all bear emphasis, because many of the unsolved mysteries of 21st-century biomedicine will require large-scale, collective action to resolve. Whether to proceed with such studies calls for collective, rather than individual, decision-making: Do we, as a group, wish to gain this knowledge or not? If so, then our collective decision whether to proceed, and on what terms, must bind everyone.

Governance "in the sense of binding collective decisions about public affairs"[197] is one of a core set of universal concepts—such as giving, lending, reciprocation, and coalition—that anthropologists find to be widely shared across many different cultures and societies, however primitive or advanced they may be. People acting in solidarity can reap benefits that autonomous individuals, acting alone, may forfeit, and human populations taken as a whole are greater than the sum of their atomistically autonomous parts. This concept was under-theorized in twentieth-century bioethics. Autonomy-based bioethics disempowers the people it seeks to protect if it precludes collective action on matters of common interest

As Charles Taylor pointed out in discussing whether it violates individual freedom for the state to install a traffic light at a frequented intersection, thus forcing people to stop: a philosopher could find a violation in this, but most people view their autonomy with a sense of proportion.[198] "[I]n such a case it is incorrect to speak of an infringement of freedom: the security and convenience of the walkers are in question, not freedom."[199] The use of data to advance precision medicine implicates patient safety, public health, and the preservation of other people's lives—not individual freedom. Bioethical principles that support a right of individuals

---

192 Faden RR, Kass NE, Goodman SN, Pronovost P, Tunis S, Beauchamp TL. *An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics*. The Hastings Center report 2013;Spec No:S16-27, at 16.

193 *Id*. at S18.

194 *Id*. at S23.

195 *Id*.

196 *Id*.

197 Stuart P. Green, *The Universal Grammar of Criminal Law: Basic Concepts of Criminal Law by George P. Fletcher*, 98 MICH. L. REV. 2104, 2112 (2000) (citing DONALD E. BROWN, HUMAN UNIVERSALS (1991)); *see also* Robin Bradley Kar, *The Deep Structure of Law and Morality*, 84 TEX. L. REV. 877, 885 (2006) (same).

198 CHARLES TAYLOR, PHILOSOPHY AND THE HUMAN SCIENCES, 2 PHILOSOPHICAL PAPERS 217-218 (1985) (noting, "In philosophical argument we might call this a restriction of freedom, but not in a serious political debate').

199 EMILIO SANTORO, AUTONOMY, FREEDOM AND RIGHTS: A CRITIQUE OF LIBERAL SUBJECTIVITY 247 (2003).

to veto the use of their data for such purposes blur the line between individual autonomy and narcissism.[200]

Our legal system traditionally employs informed consent when people are making decisions about risk to themselves,[201] but not when they make decisions about matters of public safety and welfare. Thus, informed consent is irrelevant when setting the speed limit or levying taxes. There is no opt-in that nullifies the speed limit for individuals who refuse their consent to drive 60 miles per hour." There is no opt-out that sets a general speed limit but allows determined speeders to fill out a form to be excused from it. Decisions about speed limits are confided to elected representatives and, once made, are binding on everyone. Taylor notes that people apply the concept of infringements on freedom "against a background of understanding that certain goals and activities are more significant than others."[202] There is a strong case that decisions to make data available for projects like the PMI, the Cancer Moonshot, and the learning healthcare system, while they pose some privacy risk for the individual, are mainly decisions about public safety—and this may be true regardless of whether a study constitutes "research" or "public health" under the traditional and increasingly blurred conceptions of those terms.

There is wide dissensus concerning appropriate policies for data use, and consumer-driven data commons allow people to develop approaches congenial to themselves. Each consumer-driven data commons could establish rights and duties of membership. For example, one commons group might establish a duty for its members to boycott clinical research projects that do not return results to participants, because individuals' access to their own data is fundamental to sustenance of the a consumer-driven data commons. Another might bestow privileges of membership, such as the privilege of using the collective data resource to inform the interpretation of one's own genomic test results. Those wishing to derive the benefits of the collective resource would be expected to contribute their own data in return. Bioethical concerns about coercion have enabled free-riders to enjoy the benefits of other people's research participation without themselves participating. This may have had strong ethical justification in the context of clinical research whether the burdens of participation are great, but is it still appropriate in the context of informational research? Commons-forming consumer groups would parse these ethics for themselves, perhaps engaging external ethics experts to inform their collective decisions.

Consumer-driven commons group could be organized by the members themselves, by patient advocacy groups, or by commercial entities acting as organizers and trustees to manage people's collective data according to rules the people themselves would set. Over time, some groups' rules might have wide appeal, attracting broader membership and amassing larger, more valuable data resources. Assembling and maintaining high-quality data resources is costly and requires external technical, legal, and ethics consultants. To fund these costs and add value to their collective data resources, members of a consumer-driven data commons might agree on acceptable revenue mechanisms to monetize their data assets. Here, consumer-driven commons

---

200 *See* "Definition of Narcissistic Personality Disorder" at http://www.healthyplace.com/personality-disorders/malignant-self-love/narcissistic-personality-disorder-npd-definition/ (defining narcissim as "a pattern of traits and behaviors which signify infatuation and obsession with one's self *to the exclusion of all others* and the egotistic and ruthless pursuit of one's gratification, dominance and ambition" [emphasis added]).

201 JOEL FEINBERG, HARM TO SELF.

202 Taylor, *supra* note 199.

offer a distinct advantage over traditional, data-holder-driven commons. Data holders face restrictions, such as the HITECH Act's restrictions on sales of data and allowable fees for data-related services,[203] that limit their potential revenue models. Law does not similarly restrict the right of individuals to sell or charge fees for preparation and transmittal of their own data. This difference may be crucial, if it makes consumer-driven data commons more sustainable than the data-holder-driven commons of the past.

The specific content of consumer-driven commons arrangements would be determined by collective agreement of the group members, and therefore could vary widely. The content of these agreements is a topic requiring much further debate and engagement of the scholarly community. The aggregated data resources of a commons-forming group would be a resource shared by the group members and, at a minimum, the group's agreed set of rules should address the following issues identified by Schlager and Ostrom:[204]

First, a consumer-driven data commons should have transparent rules that describe "operational-level" entitlements: What rights do members have to access their data that is in the collective data set? What duties do they have to contribute data and to update their data contributions? What uses can they personally make of the collective data set? Should the group appoint a trustee—perhaps a commercial data managaement service--to manage the day-to-day process of exercising HIPAA access rights to collect members' data from external data holders? Will it engage consultants to develop a common data model and convert members' data into a consistent, useful format? Should it appoint an ethics advisory committee to advise it on ethical issues, and what privacy and data security arrangements should it adopt? How will necessary management and consulting services be funded? Will members pay a fee for group membership, or are they willing to charge third-party data users for access to their collective data resources? Do they wish to promote certain data uses, for example, by charging academic users lower fees than they charge commercial users, or by providing cut-rate access to users who commit to place resulting discoveries in the public domain? Do they wish to encourage socially beneficial behaviors among parties who access and use their data, for example, by getting drug companies that use their data to commit to making drugs available for reasonable, cost-based fees as opposed to whatever the market of desperate patients will bear? What are the rules for new members to enter the group or for existing members to exit?

Second, the group's agreed rules should describe how it will make decisions and the "collective choice" rights that members will have to participate in those decisions. Will members vote on requests by third parties to access and use their data resources, or will they appoint an elected IRB or panel of scientific experts to assess which data requests reflect high-quality scientific uses that should be allowed? Will individual members have veto rights over certain categories of controversial data use, or will all requests for data access be decided by majority voting? What sort of data use agreements and privacy and security commitments should the group require from all entities that wish to use their data? What is the mechanism for resolving disputes within the group?

A related question, to be decided at a societal level, concerns the role of external law: Should consumer-driven data commons be subject to external regulations that set minimal substantive standards with which all consumer-driven commons must comply, or should the role of law be limited to contractual enforcement of whatever terms group members have agreed, to

[203] *See* Evans, *supra* note 110 (summarizing these deficiencies).
[204] *See supra* notes 123-125 and accompanying text.

transparent disclosure of what those terms are, and to supervision to ensure that members can freely enter and exit as promised?

Conclusion

The ultimate goal of consumer-driven data commons is to make the concept of "public trust" obsolete. Ubiquitous calls to make large-scale data systems worthy of public trust highlight that trust is something one needs when one is not control. A person traveling by commercial airliner needs to trust the pilot. When driving one's own car, trust is less necessary: one can judge for oneself whether the driver is sober and competent. Concerns about public trust recede when the public has a meaningful voice in governance. This article has sketched a mechanism for the formation of consumer-driven data commons that would allow consenting groups of individuals to assemble datasets on a meaningful scale, empowering themselves through collective action to exercise greater control over the fate of their data than individuals can achieve acting alone. It is crucial to summarize two things that consumer-driven data commons *are not*.

First, consumer-driven data commons are not a scheme of compulsory data access. Individuals would remain free to join (or not to join) one or more commons-forming consumer groups, at the individual's sole discretion. Consumer-driven data commons preserve each individual's right to consent to research uses of his or her data; however, they move the critical points in the consent process to the moments when individuals decide whether to affiliate with, not affiliate with, remain in, or exit from specific consumer-driven data commons groups. While affiliated with a commons group, the consumer would be bound by whatever data-access rules and decision-making processes its members have collectively agreed. In effect, membership in the group would be equivalent to appointing the group's decision-making body to act as one's surrogate for purposes of informed consent to uses of one's data. However, individuals would have the right to secede from a commons group—on the terms to which the individual agreed when joining the group—if its collective decisions ever became repugnant.

The second important point is that consumer-driven data commons would not necessarily embrace ethical and privacy norms that are radically different from those reflected in Common Rule and HIPAA Privacy Rule. It is conceivable that a group of commons-forming individuals might determine after careful deliberation that they *like* the norms of data access reflected in those regulations. If so, they would be free to adopt those norms as the data-access policies of their commons. They might, however, decide to elect (or otherwise choose for themselves) the IRB that will be authorized to make waiver decisions and otherwise oversee ethical and privacy protections for their group. They might adopt IRB procedures that make IRB members directly accountable to the people whose data are used, or require them to follow more rigorous due-process protections than the Common Rule provides.

Alternatively, a commons-forming group might decide that it dislikes the Common Rule and Privacy Rule. We sometimes forget that the Common Rule and Privacy Rule were not handed down on stone tablets from a source of Ultimate Ethical Truth. Rather, both regulations set *minimal* standards. The Common Rule enunciates minimal ethical protections that many federal agencies require researchers to offer to research subjects, in order to be eligible to receive federal funding. At no point were research subjects given a direct, meaningful voice in establishing those standards. It is possible that individuals, if given the opportunity to do so, might design better ethical and privacy standards, more responsive to the concerns of people whose data are used in research, yet capable of supporting a vibrant research enterprise that benefits us all. Consumer-driven data commons are a mechanism to let consumers try.

The disconnect between survey data (which suggest that 80% of Americans feel favorably about letting researchers use their data) and enrollment data (which show very few Americans actually consent for their data to be used) may be signaling that the Common Rule and Privacy Rule are not what consumers want. Each consumer-driven data commons would be free to enunciate its own ethical, privacy, and data-access policies—the terms on which its members are willing to allow their data to be used. As multiple groups enunciate their policies, there would be a "marketplace of ethical and privacy policies," which individuals could compare when deciding which consumer-driven data commons they wish to join. Markets are an excellent generator of empirical data: Consumer-driven commons that succeed in enrolling members presumably would have enunciated policies that reflect what people want; those that fail would not have done so; consumer-driven commons groups could learn from each other. Consumer-driven commons that succeed in sponsoring useful lines of research, on terms satisfactory to their members, would have successfully threaded the needle of balancing privacy and data access—a challenge that the Common Rule and Privacy Rule have chronically failed to meet. Successful consumer-driven data commons might expand and eventually become important drivers of 21st-century science.

This paper offers consumer-driven data commons not as a panacea, but as a grand experiment in democratic data self-governance, perhaps worth trying at a time when existing mechanisms of data access seem destined not to meet the challenges that lie ahead. It is fortuitous that new forms of PHD, such as data from mobile and wearable sesing devices, are generally not regulated by the Common Rule and HIPAA Privacy Rule. This regulatory gap offers an opportunity to design a new PHD privacy and access models on a blank slate, perhaps avoiding pitfalls of existing regulations. A major pitfall, till now, has been the tendency of our federal regulations to enshrine a data-holder-centric view, in which data holders assemble large-scale data resources by invoking individual consent exceptions to free data for socially beneficial research on terms, and subject to privacy and security protections, in which the data subjects—the people whose data are used—have no real voice. The goal of consumer-driven data commons is to grant people the voice they have previously been denied and, by doing so, engage citizens more actively in the task of assembling the hundred-million-person and even billion-person cohorts that 21st-century science ultimately will require.