THE RIGHT TO BENEFIT FROM BIG DATA AS A PUBLIC RESOURCE

MARY D. FAN*

The information that we reveal from interactions online and with electronic devices has massive value—for both private profit and public benefit, such as improving health, safety, and even commute times. Who owns the lucrative big data that we generate through the everyday necessity of interacting with technology? Calls for legal regulation regarding how companies use our data have spurred laws and proposals framed by the predominant lens of individual privacy and the right to control and delete data about oneself. By focusing on individual control over droplets of personal data, the major consumer privacy regimes overlook the important question of rights in the big data ocean.

This Article is the first to frame a right of the public to benefit from our consumer big data. Drawing on insights from property theory, regulatory advances, and open innovation, the Article introduces a model that permits controlled access and the use of big data for public interest purposes while protecting against privacy harms, among others. I propose defining a right of access to pooled personal data for public purposes, with sensitive information safeguarded by a controlled-access procedure akin to that used by institutional review boards in medical research today. To encourage companies to voluntarily share data for public interest purposes, the Article also proposes regulatory sandboxes and safe harbors akin to those successfully deployed in other domains, such as antitrust, financial technology, and intellectual property law.

INTRODUCTION 1439
I. PRIVACY MYOPIA: FOCUSING ON DATA DROPLETS, MISSING THE BIG DATA OCEAN 1447
A. *What the World's Strongest Consumer Privacy Protection Paradigm Misses* 1448
B. *Consumer Data Privacy in the Wild West (United States)* 1454
1. *States Lead on Data Privacy Laws* 1455

 For excellent insights and suggestions, I thank Bret Asbury, Andrew Beers, Adam Benforado, Ryan Calo, David S. Cohen, Robert Field, Cindy Fester, Alex Geisinger, Deborah S. Gordon, Nicole Iannoraone, Anil Kalhan, Amy Landers, Rachel Moran, Tejas Narechania, Elaine Sedenberg, Kate Starbird, Norman Stein, Maurice Stucke, Toshiko Takenaka, Gabriel Villareal, Tatsuhiko Yamamoto, Emily B. Zimmerman, and participants in workshops at the University of California, Berkeley Center for Long-Term Cybersecurity, the Drexel University Thomas R. Kline School of Law, Keio University, and the University of Washington's Center for an Informed Public. For outstanding editing and contributions, I am very grateful to Articles Editor Belinda Lee, who brought important expertise to this piece, and the leadership team: Joseph Krakoff, Daphne Fong, Evan A. Ringel, and Arman Cuneo.

2. *Congressional Debates and Proposals* 1459
II. ACCESS TO CONSUMER BIG DATA: THE PUBLIC'S RIGHTS IN POOLED PERSONAL DATA 1463
A. *Property Approaches to Pooled Personal Data* 1464
1. *Property by Capture: Intellectual Property Protections for Compiled Data* 1465
2. *Beyond Ownership by Capture and Personal Data as* Ferae Naturae 1468
B. *Toward a Public Right to Benefit from Our Pooled Data* 1469
1. *Access for Public Benefit Purposes* 1470
2. *Deciding What Constitutes a Public Interest Use* 1473
III. REALIZING A PUBLIC RIGHT TO BENEFIT FROM PRIVATELY HELD CONSUMER BIG DATA 1477
A. *Preventing the Tragedy of the Data Commons: Privacy and Related Harms* 1477
B. *Controlled-Access Strategies for Privacy Protection* 1479
1. *Safeguarding Our Most Sensitive Data* 1480
2. *Applying Controlled Access to Privately Held Consumer Big Data* 1483
C. *Safe Harbors and Regulatory Sandboxes for Public Interest Sharing* 1486
1. *Statutory Safe Harbors* 1487
2. *Regulatory Sandboxes* 1489
CONCLUSION 1492

INTRODUCTION

In a classic story, hungry travelers with an empty pot come upon a wary village with scant food.[1] The travelers set up their empty pot,

[1] As with many narratives in the public domain, there are many variations of the story. *See, e.g.*, MARCIA BROWN, STONE SOUP: AN OLD TALE (1947) (the Stone Soup story with hungry soldiers); JON J. MUTH, STONE SOUP (2003) (retelling the Stone Soup story with monks rather than soldiers); Sue Kimmel, *Stone Soup: A Story About Using Story for Research*, 19 SCH. LIBRARIES WORLDWIDE 1, 2 (2013) (discussing published adaptations and variations of Stone Soup); Tom Chapin, *Stone Soup*, *on* MOTHER EARTH ALBUM (Cherry Lane Music 1992) (singing a variation of Stone Soup); *cf.* James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33, 44–46 (2003) (discussing the rich productivity that has arisen from the analogous free and open-source software movements and decrying the privatization of the "commons of the mind" and resulting chilling of such creative productivity).

light a fire, and select stones to boil.[2] Enticed by the idea that one can make food from stones, villagers emerge, intrigued, and begin to contribute their scarce provisions—a carrot here, a potato there, some herbs, and so forth, until a Stone Soup emerges, enough to feed the whole group. A bounty with the potential to benefit the larger group emerges from amassing seemingly small individual contributions.[3]

But who owns the soup? Who has the right to use (or here, derive nourishment from) it? Would it be fair if the travelers with the pot—the infrastructure to cook the soup—and the creative spark to entice individual contributions claimed sole rights to control and use the soup? And to get more sophisticated, should different levels of contributors get different bundles of rights to the soup?

We have some intuitive answers to the Stone Soup metaphor that has persisted across centuries and cultures.[4] The usual outcome of the traditional story, after all, is that the group shares in the benefits—not that the clever providers of the pot and enticing idea get it all.

Moving from the timeless to the timely questions of our era, we contribute data to a massive, growing pool every time we use the Internet or the Internet-of-things—the billions of smart devices like Amazon's Echo, Apple Watches, or Fitbits that are connected to the Internet and sharing our data.[5] In 2013, Norwegian research group SINTEF reported that ninety percent of the world's data had been

---

[2] This retelling, in broad strokes, is a composite of the many variations of the story discussed *supra* note 1.

[3] The analogy, of course, is the power gained from pooling data from individual information or cases. *See, e.g.*, MARY D. FAN, CAMERA POWER: PROOF, POLICING, PRIVACY, AND AUDIOVISUAL BIG DATA 117–21 (2019) [hereinafter FAN, CAMERA POWER] (discussing the predictive power gained by pooling together many small events and applying advanced analytics); Andrew Manuel Crespo, *Systemic Facts: Toward Institutional Awareness in Criminal Courts*, 129 HARV. L. REV. 2049, 2051 (2016) (discussing the insights to be gained from examining patterns across cases, rather than myopically constraining focus on the individual case); EUR. COMM'N, STUDY ON BIG DATA IN PUBLIC HEALTH, TELEMEDICINE, AND HEALTHCARE: FINAL REPORT 46–54 (2016), https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf (reporting findings about how big data aggregation and analytics have led to improvements and advances in public health and medicine, and recommending strategies to facilitate big data pooling and access).

[4] For a genealogy of how the Stone Soup story has spread from the 1700s to our time and across cultures, see William Rubel, *History of the Stone Soup Folktale from 1720 to Now*, STONESOUP (Sept. 2015), https://stonesoup.com/about-the-childrens-art-foundation-and-stone-soup-magazine/history-of-the-stone-soup-story-from-1720-to-now.

[5] *See, e.g.*, Stacy-Ann Elvy, *Commodifying Consumer Data in the Era of the Internet of Things*, 59 B.C. L. REV. 423, 426–28 (2018) (discussing the wealth of information amassing from users accessing the Internet-of-things); Xuan-Thao N. Nguyen, *Collateralizing Privacy*, 78 TUL. L. REV. 553, 564–67 (2004) (discussing the aggregation of consumer data into valuable datasets and databases).

generated in the preceding two years.[6] By 2020, the volume was projected to grow to forty-four zettabytes—forty times more bytes than the number of stars in the observable universe—according to the World Economic Forum.[7] The vast volumes of data facilitate the training and deployment of advanced analytical techniques that power artificial intelligence, including machine learning.[8]

Pooled together and processed using advanced analytics, the data and findings could be powerfully deployed for both private profit and public benefit—if the data were accessible for use. Vast volumes of data are valuable for advanced analytical techniques, such as machine learning, to address a range of goals, such as predicting and preventing disease, understanding the spread of misinformation, and improving traffic congestion.[9] The data also can be used for commercial advantage, to expand markets, tailor offerings, sell more products, or even exploit cognitive weaknesses to get people to overspend or waive rights.[10]

---

[6] *See* SINTEF, *Big Data, for Better or Worse: 90% of World's Data Generated over Last Two Years*, ScienceDaily (May 22, 2013), https://www.sciencedaily.com/releases/2013/05/130522085217.htm.

[7] Jeff Desjardins, *How Much Data Is Generated Each Day?*, World Econ. F. (Apr. 17, 2019), https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f.

[8] For an overview of machine learning techniques see, for example, Ethem Alpaydin, Machine Learning: The New AI (2016). For an accessible overview of the relevant concepts aimed at attorneys and non-technical readers, see Ai Deng, *An Antitrust Lawyer's Guide to Machine Learning*, 32 Antitrust 82, 82 (2018).

[9] *See, e.g.*, Kyle Wiggers, *Google Cloud Releases Covid-19 Data Sets to Foster Coronavirus-Fighting AI Models*, VentureBeat (Mar. 30, 2020, 9:39 AM), https://venturebeat.com/2020/03/30/google-launches-covid-19-public-datasets-program-to-foster-coronavirus-fighting-ai-models (reporting on free datasets and analytical programs that Google Cloud is releasing to help train machine learning models to help prevent the spread of COVID-19); Gary King & Nathaniel Persily, *Unprecedented Facebook URLs Dataset Now Available for Academic Research Through Social Science One*, Soc. Sci. One (Feb. 13, 2020) [hereinafter King & Persily, *Facebook URLs Dataset*], https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one (announcing the release of "one of the largest social science datasets ever constructed" in collaboration with Facebook to study elections and the spread of information and misinformation); Nicole Dungca, *In First, Uber to Share Ride Data with Boston*, Bos. Globe (Jan. 14, 2015), https://www.bostonglobe.com/business/2015/01/13/uber-share-ridership-data-with-boston/4Klo40KZREtQ7jkoaZjoNN/story.html (reporting that Uber is sharing ride data information with city officials, starting with Boston, in an effort to study strategies to reduce traffic congestion); Alyssa Newcomb, *Why Uber Is Sharing Ride Data with the City of Boston*, ABC News (Jan. 13, 2015, 9:16 AM), https://abcnews.go.com/Technology/uber-sharing-ride-data-city-boston/story?id=28188898 (reporting that city officials will use Uber shared data to study ways to reduce traffic congestion).

[10] *See, e.g.*, Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Maer, Marshini Chetty & Arvind Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 Proc. ACM Hum.-Comput. Interact., Nov.

Access to big data as a public resource is important to address pressing questions that cannot be answered without the data and to correct imbalances in access to information and control over knowledge. For example, are there patterns in how major social media companies filter, censor, or promote the news and content that people receive by ideology, race, and other potentially invidious factors?[11] How do data and information flows influence the rise of violent extremism?[12] How might our data be used against us to discriminate in the provision of services, advertisements, special deals, price-setting, and access to opportunities?[13] Is there racial profiling in the algorithms that determine whether and what content reaches

2019, art. 81 (presenting findings of dark patterns, user interfaces designed to trick and confuse users, used in websites targeting consumers); Jamie Luguri & Lior Jacob Strahilevitz, *Shining a Light on Dark Patterns*, 13 J. LEGAL ANALYSIS 43, 81–82 (2021) (discussing the proliferations of dark patterns in user interface designs and reporting the results of an experiment that shows the power of dark patterns to manipulate consumers); Christopher Bosch, Benjamin Erb, Frank Kargl, Henning Kopp & Stefan Pfattheicher, *Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns*, 2016 PROC. PRIV. ENHANCING TECH 237 (presenting examples of dark patterns in design).

[11] *See, e.g.*, Shannon Bond, *Facebook and Twitter Limit Sharing New York Post Story About Joe Biden*, NPR (Oct. 14, 2020, 9:14 PM), https://www.npr.org/2020/10/14/923766097/facebook-and-twitter-limit-sharing-new-york-post-story-about-joe-biden (discussing controversy over the decision by Facebook and Twitter to limit sharing of a media story and the lack of clarity or transparency over how algorithms restrict or allow sharing); Bianca Bruno, *Google & YouTube Accused of Racial Profiling*, COURTHOUSE NEWS SVC. (June 18, 2020), https://www.courthousenews.com/google-youtube-accused-of-racial-profiling (reporting on a federal class action alleging racial profiling in how algorithms filter and censor content); Gail Sullivan, *How Facebook and Twitter Control What You See About Ferguson*, WASH. POST (Aug. 19, 2014, 3:43 AM), https://www.washingtonpost.com/news/morning-mix/wp/2014/08/19/how-facebook-and-twitter-control-what-you-see-about-ferguson (discussing concerns over how information regarding fiercely partisan events like the Ferguson protests over policing is disseminated or shared).

[12] *See, e.g.*, INES VON BEHR, ANAÏS REDING, CHARLIE EDWARDS & LUKE GRIBBON, RADICALISATION IN THE DIGITAL ERA iii (2013), https://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR453/RAND_RR453.pdf (discussing the need for empirical research on radicalization and the extreme difficulty in getting access to data); Emerson T. Brooking & P.W. Singer, *War Goes Viral: How Social Media Is Being Weaponized Around the World*, ATLANTIC (Nov. 2016), https://www.theatlantic.com/magazine/archive/2016/11/war-goes-viral/501125 (discussing new dilemmas raised by the convergence of physical violence, information warfare, and online speech for companies and how platforms can be used in such warfare).

[13] *See, e.g.*, Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMM. ACM 44, 46–53 (2013) (discussing racial profiling and biases in advertising content, such as using advertisements suggesting arrest for persons with Black-identified names); Nathan Newman, Comment on FTC Big Data: A Tool for Inclusion or Exclusion? Workshop: How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population (Aug. 15, 2014), https://www.ftc.gov/system/files/documents/public_comments/2014/08/00015-92370.pdf (collecting concerns over the use and misuse of consumer data to engage in price discrimination, information content delivery, and the availability or denial of opportunities).

viewers?[14] How do internet bots and trolls spread viral patterns of misinformation and potentially even influence vital decisions such as presidential elections or COVID-19 vaccine refusal?[15] There are major potential distributional consequences for who or what entities get to benefit from our pooled personal information—and who is locked out.[16] Big data access also is important to better inform law and policy. Governmental institutions, such as administrative agencies and courts, increasingly seek more rigorous empirical information to guide decisions and policies, but key data may not be available or may be privately controlled.[17]

Who has the right to control the data is a major issue that legislatures are beginning to confront—usually from the perspective of individual privacy.[18] The individual privacy protection framework is important but does not address the larger overarching question about who gets to use and benefit from the valuable amassed data, which are currently often held by private companies. Many companies have adopted restrictive policies on voluntary sharing of information that risk blocking socially valuable research or potentially biasing studies

---

[14] *See, e.g.*, Class Action Complaint for Declaratory Judgment, Restitution and Damages, Newman v. Google, No. 20-cv-04011 (N.D. Cal. June 16, 2020), ECF No. 1 (alleging that Google and YouTube algorithms discriminate against Black content creators and consumers).

[15] *See, e.g.*, LEE RAINIE, JANNA ANDERSON & JONATHAN ALBRIGHT, PEW RSCH. CTR., THE FUTURE OF FREE SPEECH, TROLLS, ANONYMITY, AND FAKE NEWS ONLINE 2–5 (2017), https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2017/03/PI_2017.03.29_Social-Climate_FINAL.pdf (discussing rising questions and concerns over internet bots, trolls, and manipulative behaviors affecting important issues such as elections).

[16] *See, e.g.*, EMMA ROOKSBY & JOHN WECKERT, INFORMATION TECHNOLOGY AND SOCIAL JUSTICE, at vi–vii (2007) (discussing the distributive justice problems raised by severe inequalities in access to information); LUKE W. COLE & SHEILA R. FOSTER, FROM THE GROUND UP: ENVIRONMENTAL RACISM AND THE RISE OF THE ENVIRONMENTAL JUSTICE MOVEMENT 109 (2001) (discussing how lack of information and access to knowledge creates systematic inequality, "severe barriers to participation in a pluralistic process," and unequally distributes political power).

[17] *See, e.g.*, Christian Leuz, *Evidence-Based Policy-Making: Promise, Challenges, and Opportunities for Accounting and Financial Markets Research*, 48 ACCTG. & BUS. RSCH. 582, 583, 589 (2018) (discussing the rise of evidence-based policy-making and how lack of data is the biggest challenge to meeting calls for more rigorous empirical evidence); RODNEY A. SMOLLA, 2 RIGHTS AND LIABILITIES IN MEDIA CONTENT § 11:17 (2d ed. 2021) (noting that "[e]mpirical evidence or the lack of it will often play an important role in the application" of the Supreme Court's standard for regulation of commercial speech and collecting cases by courts expressing concern over the lack of empirical data and the need for more rigorous evidence regarding the rationales of governmental bodies for regulating commercial speech and the impact of such regulation).

[18] For a discussion of recent legislative proposals and laws at the federal and state levels, and in the European Union, see *infra* Sections I.A–B.

and results.[19] The lens of privacy protection leads to laws and regulations focused on an individual's control over personal data.[20] Yet an individual's data point is just a drop in the sea of what is valuable, powerful, possibly dangerous, and potentially beneficial about pooled data points.[21]

This Article fills the gap, framing and making the case for a right to use and benefit from our pooled consumer data. Commercial entities provide the infrastructure to attract and collect our data, but the pooled data has a public resource nature in several respects.[22] First, we are all contributing to the generation of the resource through our activities and interactions with online platforms and devices. Second, private control and ownership does not put the resource of our pooled data to its socially optimal use—and indeed may be conducive to nefarious uses, as the current concerns over dark patterns and political manipulation show.[23] Third, nonexclusive access could put the resource to its socially optimal use, because such access would enhance healthy commerce and prevent private capture that stunts realization of the full utility of our pooled electronic data.[24]

Of course, our collective electronic data is a more complicated form of resource than traditional commons or group property such as coastal fisheries, roadways, common fields, or waterway access. Because data are intangible and reproducible, the danger of mismanagement and uncontrolled access is not depletion, but rather privacy and misuse harms.[25] Further, commercial entities provide the infra-

---

[19] *See, e.g.*, David M.J. Lazer et al., *Computational Social Science: Obstacles and Opportunities*, 369 SCIENCE 1060, 1060 (2020) (noting "many companies have been steadily cutting back data that can be pulled from their platforms" because of regulations or scandals such as the Cambridge Analytica scandal, resulting in the closing of "avenues of potentially valuable research" and a voluntary, arbitrary data-sharing system that is "intrinsically unreliable and potentially biased in the science it produces").

[20] For a discussion of provisions reflecting this focus, see *infra* Sections I.A–B.

[21] *See, e.g.*, Alex Romanov, *How Much or Too Little? Assigning a Dollar Value to Big Data*, VENTUREBEAT (Nov. 5, 2013, 2:30 AM), https://venturebeat.com/2013/11/05/data-worth (explaining how the "small . . . value[] assigned to each individual data point—even as little as a few cents—multiplied by large numbers . . . equals the potential for significant return on investment").

[22] For a discussion of characteristics of certain properties or resources that make them more akin to public property rather than better kept under private ownership, see, for example, Carol Rose, *The Comedy of the Commons: Custom, Commerce, and Inherently Public Property*, 53 U. CHI. L. REV. 711, 713–23, 774–77 (1986).

[23] *See, e.g.*, Tom McKay, *Senators Introduce Bill to Stop 'Dark Patterns' Huge Platforms Use to Trick Users*, GIZMODO (Apr. 9, 2019, 9:30 PM), https://gizmodo.com/senators-introduce-bill-to-stop-dark-patterns-huge-plat-1833929276 (discussing legislative proposal to address concerns); Mathur et al., *supra* note 10, at 244 (cataloguing examples); Bosch et al., *supra* note 10 (same).

[24] *See* discussion *infra* Part II.

[25] *See* discussion *infra* Part III.

structure and impetus for us to provide data, have default control over access, and enjoy potential copyright and trade secret protections for the data compilations.[26] But while there are limits to the analogy, there also are important lessons to gain.

Recognizing that our pooled personal data has a public-resource nature opens our vision to the great promise and peril of shared access and important lessons on governance. This right of access to our pooled data for research in the public interest can build off the current baseline of property protections in data compilations that companies have invested in creating. The Article also proposes a controlled-access approach to maximize the public benefit potential of shared use of our pooled data while protecting people from privacy harms, among others.

The Article illuminates how at least two areas of law designed with for-profit transactions as the paradigm have the side effect of disincentivizing data sharing for nonprofit public interest purposes. The two major privacy regimes in force today, the European Union's General Data Privacy Regulation (GDPR) and the California Consumer Privacy Act (CCPA), largely overlook the importance of access to our pooled consumer big data for nonprofit and public interest purposes in their focus on business use of personal information.[27] More problematically, these regimes potentially chill sharing of the data to benefit the public because of stringent requirements and penalties for privacy violations.[28] Second, trade secret law's protections depend on keeping valuable information secret.[29] The Article offers proposals to reform these unintended adverse consequences.[30]

Now is an opportune time to investigate and frame the right of the public to benefit from our data collected by private entities. Recently proposed legislation in the United States and European Union are making early forays into crafting limited data-sharing obligations. Introduced in the House of Representatives in late May 2021, the U.S. Social Media Disclosure and Transparency of Advertisements Act would require digital platform giants with more than 100 million monthly active users to share targeted advertisement data with aca-

---

[26] *See, e.g.*, Mark Birkin, *Spatial Data Analytics of Mobility with Consumer Data*, 76 J. TRANSP. GEOGRAPHY 245, 250–51 (2019) ("Since consumer data arise as the product of a commercial transaction between an organization and its customers the data are typically owned and controlled by business organisations which are external to the academic sector.").

[27] *See* discussion *infra* Sections I.A–B.

[28] *See* discussion *infra* notes 65–85 and accompanying text.

[29] *See* discussion *infra* Section II.A.

[30] *See* discussion *infra* Sections II.B, III.B.

demic researchers.[31] In the European Union, the European Parliament and member states are considering the proposed Digital Services Act, which obligates "very large online platforms," with average monthly active user bases of 45 million or more, to give data access to "vetted researchers . . . for the sole purpose" of researching "systemic risks" posed by use of the platforms.[32] European Union members also are considering another proposal, the Data Governance Act, which contains provisions to facilitate the development of "data altruism organizations" that serve as intermediaries to make available pooled data consensually shared by people for "altruistic purposes" such as scientific research or improving public services.[33] While limited in scope and subject-matter, these proposals are valuable early efforts towards the larger issue this Article addresses, of access to privately controlled consumer data.

The Article proceeds in three parts. Part I discusses how concerns over the ways in which companies control and use our data has led to a proliferation of legislative proposals and laws attempting to regulate consumer data. This Part argues that the predominant focus on individual privacy spurring this first wave of laws is important but also myopic, failing to address the larger overarching question regarding the right of the public to benefit from the collection of our data by private entities.

Part II advances the case for recognizing the shared public resource nature of our amassed data, drawing on property theory and insights from open innovation movements. This Part argues that while private entities may provide the infrastructure and enticements to amass the information, there remains a right of the public to benefit from our individual contributions. Moreover, recognizing the public's right to benefit is in the interest of private entities that currently control our amassed consumer data as well as in the interest of the public.

Part III addresses important concerns that expanding access to researchers whose work provides a public benefit raises privacy concerns and may limit incentives for private investments in cultivating and safeguarding the data. Drawing insights from commons-

---

[31] Social Media Disclosure and Transparency of Advertisements Act of 2021 (Social Media DATA Act), H.R. 3451, 117th Cong. (introduced May 20, 2021).

[32] *Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, arts. 25(1), 26(1), 31, COM (2020) 825 final (Dec. 15, 2020), https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN.

[33] *Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*, arts. 2(10), 15–22, COM (2020) 767 final (Nov. 25, 2020), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767.

governance theory and time-tested protections for human subjects in research, this Part proposes a controlled-access model for providing public benefits from use and access while reducing the risk of privacy harms, among others. A right to use our amassed data for public benefit does not mean that everyone gets to access and use the sensitive data. Rather, access and use should be governed by the well-established principles and safeguards that have led to lifesaving research using some of our most sensitive, protected, and private data—health information. Insights and experiences from the regulation of human subjects research show the feasibility of recognizing a right of access to pooled personal data for research in the public interest while addressing privacy and related concerns.

I
Privacy Myopia: Focusing on Data Droplets, Missing the Big Data Ocean

Public discontent over control, access, and use of personal data is brewing—and with it the seeds of democratic change via legislation. A recent nonpartisan Pew Research Center survey of a random sample of more than 10,000 Americans found that 79% are concerned over how companies use collected personal data.[34] Americans overwhelmingly (72%) believe their consumer data is not being used to benefit them.[35] Rather, Americans are widely aware (77%) that companies use the data to profile them, engage in targeted advertising, and assess their riskiness as customers.[36] Consumers in the United Kingdom, Germany, and India—who, along with Americans, represent the largest portion of online users—have similarly widespread concern over how companies use their data, according to a 2014 survey.[37] The survey found that 97% of respondents were concerned about misuse of their data.[38]

---

[34] Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner, Americans and Privacy: Concerned, Confused, and Feeling Lack of Control over Their Personal Information 8 (2019), https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information. For information on the survey's size, see *The American Trends Panel*, Pew Rsch. Ctr. (2019) [hereinafter *American Trends Survey Methodology*], https://www.pewresearch.org/methods/u-s-survey-research/american-trends-panel (reporting 10,000-plus panelists).

[35] Auxier et al., *supra* note 34, at 7.

[36] *Id.*

[37] Timothy Morey, Theodore Forbath & Allison Schoop, *Customer Data: Designing for Transparency and Trust*, Harv. Bus. Rev. (May 2015), https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust.

[38] *Id.*

To address the exponential increase of personal data that companies are amassing and concerns over how it is used, laws and legislative proposals are emerging in states, the U.S. Congress, and the European Union. As discussed below, most of the focus in this first wave of legislation is on personal control over individual data and privacy protection. While the privacy-oriented laws and proposals address one facet of the rising concern over the use of amassed data, they leave unanswered other important access and distributional concerns about who gets to use and benefit from the pooled data collected from the public and held in privatized silos. The legislative fomentation around consumer data privacy presents fertile ground for fresh approaches to the control, access, and use of our pooled personal data—but also reveals gaps in regulation, as discussed below.

A. What the World's Strongest Consumer Privacy Protection Paradigm Misses

The European Union (EU) was the early major mover in defining individual rights in personal data and the obligations of companies by enacting the GDPR, which took effect on May 25, 2018.[39] The GDPR offers "data subjects" a data bill of rights.[40] The GDPR also requires data "controller[s]" to implement "data protection by design and by default," secure and protect data, and conduct "data protection impact assessment[s]," among other obligations.[41] The additional protections expand on the GDPR's predecessor in regulation, the 1995 Data Protection Directive.[42]

Professors Paul Schwartz and Dan Solove have observed a difference in cultural orientations toward privacy rights in the European Union and the United States: "EU law views privacy as a fundamental right, while U.S. law considers it one interest that is balanced against others."[43] The GDPR's approach reflects this orientation, operationalizing the commitment to the fundamental rights to privacy and family life expressed in the Charter of the Fundamental Rights of the European Union.[44] The provisions on data user rights and data con-

---

[39] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 2016 O.J. (L 119) [hereinafter GDPR].

[40] *See id.* arts. 12–23 (rights of the data subject).

[41] *See id.* arts. 24–43 (obligations of controllers and processors).

[42] Council Directive 95/46, 1995 O.J. (L 281) 31 (EC).

[43] Paul M. Schwartz & Daniel J. Solove, *Reconciling Personal Information in the United States and European Union*, 102 CALIF. L. REV. 877, 880 (2014).

[44] Charter of Fundamental Rights of the European Union, arts. 7–8, 2010 O.J. (C 83) 393; *see also, e.g.*, Mira Burri, *The Governance of Data and Data Flows in Trade*

troller obligations are so strong that the GDPR "set off a wave of unease in many data-intensive industries, including health research, sparking fears that it would effectively curtail their activities."[45] While the paradigmatic concern may be unscrupulous exploitation and sale of personal information for profit, the GDPR's wide sweep even has a potential chilling effect on data-intensive nonprofit research that offers important public benefits such as advancing the treatment of disease and preventing deaths.[46]

The GDPR specifies three main types of individual rights.[47] The first category of rights is about access to information and notice. Data subjects have the right to transparent, intelligible communication about how their information is being processed and how data controllers are responding to their requests to exercise their personal rights in their data.[48] People also have the right to information about the data controller, their rights, and whether and how their data will be used, stored, and processed, including access to a copy of that personal data.[49]

The second cluster of rights pertains to control over one's personal data. The GDPR confers a right to rectification of inaccurate personal data, including supplementation of incomplete information.[50] People also have the right to erasure of personal data under certain conditions, such as withdrawal of consent or if the data are no longer

---

Agreements: The Pitfalls of Legal Adaptation, 51 U.C. DAVIS L. REV. 65, 89 (2017) ("[T]he GDPR is a piece of EU legislation that is meant to provide for very high standards of protection of personal data as an expression of the fundamental rights of EU citizens to privacy and family life, as embedded in the Charter of Fundamental Rights of the EU.").

45 Mark Phillips & Bartha M. Knoppers, *Whose Commons? Data Protection as a Legal Limit of Open Science*, 47 J.L. MED. & ETHICS 106, 107 (2019).

46 *See* Jasper Bovenberg, David Peloquin, Barbara Bierer, Mark Barnes & Bartha Maria Knoppers, *How to Fix the GDPR's Frustration of Global Biomedical Research*, 370 SCIENCE 40, 41–42 (2020) (discussing the barriers the GDPR imposes to data sharing for lifesaving biomedical research, such as addressing COVID-19); David Peloquin, Michael DiMaio, Barbara Bierer & Mark Barnes, *Disruptive and Avoidable: GDPR Challenges to Secondary Research Uses of Data*, 28 EUROPEAN J. HUM. GENETICS 697, 697–700 (2020) (discussing the problems posed by the GDPR for secondary research beyond the initially approved study, and for the development of biobanks and databanks to facilitate biomedical research); Niamh Clarke, Gillian Vale, Emer P. Reeves, Mary Kirwan, David Smith, Michael Farrell, Gerard Hurl & Noel G. McElvaney, *GDPR: An Impediment to Research?*, 188 IRISH J. MED. SCI. 1129, 1129 (2019) ("At the very least, the regulations, as applied in Ireland, will place a significant extra burden of work on Ireland's clinical researchers and at their worst will force individuals and institutions out of the clinical research field . . . .").

47 GDPR, *supra* note 39, arts. 12–22; *see also id.* art. 23 (setting forth limitations to the rights).

48 *Id.* art. 12.

49 *Id*. arts. 13–15.

50 *Id*. art. 16.

necessary for the purposes collected.[51] This implements a variant of the theoretical "right to be forgotten" that has occasioned much scholarly commentary and advocacy over the last decade and has found expression in various national and international laws, but not in U.S. law.[52] There also is the right to restrict processing of one's data under certain circumstances, such as while one is seeking to rectify errors.[53] To fully realize these rights of control, a person, dubbed a data subject, also has the right to have the data controller notify entities that receive the subject's data that errors have been rectified, data processing has been restricted, or the data must be erased.[54] The GDPR also confers a right to the portability of one's data, including the right to receive one's data "in a structured, commonly used and machine readable format" and to transmit that information to another data controller.[55] The regulations also specify a right to object that the controller's use of a person's data is unlawful.[56]

Third, the GDPR addresses some of the growing concerns about profiling and automated decisionmaking facilitated by big data analytics.[57] Commentators have expressed concern over the rise of decisionmaking by machine learning algorithms, potentially trained on biased datasets, to make impactful decisions about people's lives, such

---

[51] *Id*. art. 17.

[52] *Compare* Mosha v. Yandex, Inc., No. 18-CV-5444, 2019 WL 5595037, at *2 (S.D.N.Y. Oct. 30, 2019) (discussing Russian law's "'[r]ight to be forgotten' . . . which allows an individual to request that search engine operators . . . remove links that are incorrect or outdated"), *and* Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 43 (2013) (proposing a framework for online obscurity and arguing it is less costly to operationalize than the right to be forgotten), *and* Jeffrey Rosen, *The Right to Be Forgotten*, 64 STAN. L. REV. ONLINE 88, 88 (2012) ("In Europe, the . . . roots of the right to be forgotten can be found in French law, which recognizes *le droit à l'oubli* . . . a right that allows a convicted criminal who has . . . been rehabilitated to object to the publication of the facts of his conviction and incarceration."), *with* Garcia v. Google, Inc., 786 F.3d 733, 745–46 (9th Cir. 2015) ("[A] 'right to be forgotten,' although recently affirmed by the Court of Justice for the European Union, is not recognized in the United States." (citing Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos, ECLI:EU:C:2014:616 (May 13, 2014) (obligating Google to respond to requests to remove personal information))), *and* Yeager v. Innovus Pharms., Inc., No. 18-cv-397, 2019 WL 447743, at *7 n.6 ("But no 'right to be forgotten' exists under United States law.").

[53] GDPR, *supra* note 39, art. 18.

[54] *Id*. art. 19.

[55] *Id*. art. 20.

[56] *Id*. art. 21.

[57] *Id*. art. 22; *see also, e.g.*, Article 29 Data Prot. Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, WP251rev.01, at 5 (Feb. 6, 2018) [hereinafter Data Prot. Working Party, *Guidelines*] (noting that advances in technology, "big data analytics, artificial intelligence and machine learning have made it easier to create profiles and make automated decisions with the potential to significantly impact individuals' rights and freedoms").

as eligibility for jobs, loans, housing and school admissions.[58] The GDPR gives people a right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[59] Examples of such decisions producing legal effects or significant impact include automatic refusal on a credit application or e-recruiting automatic screens without human involvement.[60]

The GDPR recognizes exceptions to the right against automated decisionmaking, such as consent or contractual necessity.[61] The GDPR also permits a member state or a European Union law to authorize automated decisionmaking, if the authority "lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests."[62] The GDPR's provisions on automated decisionmaking do not comprehensively address the full range of objections raised to the practice, such as the "black box" nontransparent nature of algorithms used to make decisions.[63] The GDPR is, however, one of the first enacted major efforts to regulate automated decisionmaking.[64]

---

[58] For some of the rich literature on concerns over profiling and automated decisionmaking, see, for example, VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT (2017); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016). *See also, e.g.*, Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 636–54 (2020) (contrasting machine action and human decisionmaking and comparing concerns about both); Peter K. Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 FLA. L. REV. 331, 343–59 (2020) (cataloguing problems with algorithmic decisionmaking); Céline Castets-Renard, *Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making*, 30 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 91, 99 (2019) ("[B]ig data analytics, artificial intelligence, and machine learning's capabilities have significantly facilitated the creation of profiles and automated decisions with the potential to impact individual's rights and freedoms—especially when the decision concerns an application to enter a school or to obtain social benefits."); Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 505–06 ("Automated decision-making, profiling, and related machine-learning techniques pose new opportunities for privacy-invasive, discriminatory, and biased decision-making based on inferential analytics.").

[59] GDPR, *supra* note 39, art. 22(1).

[60] *See id*. recital 71; Data Prot. Working Party, *Guidelines*, *supra* note 57, at 21–22.

[61] GDPR, *supra* note 39, art. 22(2).

[62] *Id.*

[63] *See* Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841, 863 (2018) ("[T]he GDPR does not appear to require opening the 'black box' to explain the internal logic of the decision-making system to data subjects.").

[64] *Id.* at 842.

The rights and obligations in the GDPR are backed by potentially substantial legal sanctions, including potentially hefty fines.[65] Data users also may sue for damages.[66] The tariff of fines vary by type of violation.[67] At the high end, infringement of the basic legal conditions for data processing, such as satisfying the conditions for obtaining consent, can result in "administrative fines up to 20 000 000 EUR [approximately U.S. $23.5 million], or . . . up to 4 % of the total worldwide annual turnover of the preceding financial year, whichever is higher."[68] The fines are so substantial that many U.S. companies, accustomed to operating under a much more permissive balance of business interests and privacy, lack sufficient coverage to pay if sanctioned.[69]

Framing the rights of data subjects through the paramount lens of individual privacy results in a focus on an individual's control over personal data, rather than the public's interest in the aggregated personal data. The right of access, for example, is couched in terms of the singular "data subject" with respect to one's individual data, rather than vested in all data subjects for a public interest purpose such as scientific studies to improve health or safety.[70] The concession to public interest is mainly in the form of limited authorization to derogate from some of the privacy rights in the GDPR.[71]

Member states or the European Union may provide for derogations from the rights of access, rectification, restriction of processing, and objection if they "are likely to render impossible or seriously impair the achievement" of processing for scientific or statistical purposes.[72] The authorization to derogate is subject to safeguards such as pseudonymization, to respect the principle of data minimization.[73] Data minimization means ensuring that data collection, use, and storage are limited to the amount necessary for the purposes gathered.[74]

---

65 *See* GDPR, *supra* note 39, ch. VIII.

66 *Id.* art. 82.

67 *See, e.g.*, *id.* art. 83 (setting forth administrative fines by type of violation); Lukas Feiler, *Takeaways from the First GDPR Fines*, CYBERSPACE LAW., Jan.–Feb. 2018 (reporting on early three cases of actual fines imposed for data privacy breaches ranging from 4,800 euros to 400,000 euros).

68 GDPR, *supra* note 39, art. 83(5).

69 *See* Henry Kenyon, *U.S. Firms May Be Lacking in Cyber Insurance Coverage Against GDPR Fines*, CQ ROLL CALL, Sept. 21, 2018.

70 GDPR, *supra* note 39, art. 15.

71 *Id.* art. 89(1).

72 *Id.* art. 89(2).

73 *Id.* art. 89(1).

74 *Id.* art. 89.

The GDPR's underdeveloped distinction between commercial use of consumer data and nonprofit use for purposes such as advancing public health have raised concerns among researchers about potentially severe burdens.[75] Private companies such as Facebook have invoked the GDPR's threat of multimillion dollar sanctions as a reason not to share important data with researchers addressing issues such as the transmission of misinformation.[76] Ultimately, Facebook recently released one of the largest social science databases ever assembled but with details obscured using differential privacy techniques that perturb the data by introducing noise and censoring.[77] These techniques render the ability to draw research conclusions complicated, problematic, and in the case of study of smaller groups, impossible.[78]

Some obstacles to scientific research posed by the GDPR, such as questions about the specificity of consent that data controllers must obtain, have been ameliorated by later clarifications in recitals to the GDPR.[79] In 2020, the European Data Protection Supervisor noted the vital role of researchers in a democracy and how privacy protections can be used as cover for corporate secrecy to stymie data access.[80]

---

[75] *See* Edward S. Dove, David Townend & Bartha M. Knoppers, *Data Protection and Consent to Biomedical Research: A Step Forward?*, 384 LANCET 855, 855 (2014) (noting potentially significant burdens on large-scale research entities); Phillips & Knoppers, *supra* note 45, at 107 (describing "a wave of unease in many data-intensive industries").

[76] *See* King & Persily, *Facebook URLs Dataset*, *supra* note 9 ("The greatest barrier we have faced [in obtaining data from Facebook for social science research] concerned Facebook's interpretation of the relevant privacy restrictions contained in the General Data Protection Regulation (GDPR) . . . . They sometimes take the position that those restrictions prevent researchers from analyzing individual level data, even if de-identified or aggregated."); *see also* EUR. DATA PROT. SUPERVISOR, A PRELIMINARY OPINION ON DATA PROTECTION AND SCIENTIFIC RESEARCH 2 (2020), https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf ("Researchers operating within ethical governance frameworks should therefore be able to access necessary . . . data, with a valid legal basis and subject to the principle of proportionality and appropriate safeguards.").

[77] *See* King & Persily, *Facebook URLs Dataset*, *supra* note 9; *see also* Fang Liu, *A Statistical Overview on Data Privacy*, 34 NOTRE DAME J.L. ETHICS & PUB. POL'Y 477, 482–84 (2020) (discussing how differential privacy techniques generally operate); Michael Hawes, Senior Advisor for Data Access & Priv. Rsch. & Methodology Directorate, U.S. Census Bureau, American Statistical Association Webinar: Differential Privacy and the 2020 Decennial Census (Jan. 28, 2020), https://zenodo.org/record/4122103/files/Privacy_webinar_1-28-2020.pdf (last visited Sept. 13, 2021) (discussing differential privacy techniques in the context of highly protected U.S. Census data).

[78] *See* King & Persily, *Facebook URLs Dataset*, *supra* note 9.

[79] *See* Phillips & Knoppers, *supra* note 45, at 108 (describing the changes incorporated in Recital 33 to the GDPR); Edward S. Dove, *The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era*, 46 J.L. MED. & ETHICS 1013, 1013–15 (2018).

[80] European Data Prot. Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research*, at 9 (Jan. 6, 2020), https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf.

Such clarifications and acknowledgments have quelled some of the initial concerns, generating greater acceptance of the GDPR as "a well-drafted piece of legislation that raises the standards of data protection globally."[81] But there remain difficulties in interpretation and application because the GDPR's "default lens tends to focus on relationships between private sector companies and their customers."[82] Researchers continue to express concerns that digital platforms are limiting data access to mitigate the risk of liability, reducing researchers to alternative data-scavenging strategies such as building tools for unsanctioned data-scraping.[83]

The GDPR has been influential in setting standards for data protection beyond its territorial scope, as companies streamline operations across borders and nations wishing to do business in Europe adopt equivalent protections.[84] While the GDPR is a progressive—some might argue aggressive—step in data protection law, its vision remains limited by the frame of individual privacy rights and focus on the relationship between consumers and private companies using data for profitmaking ends. More work remains to be done in resolving open questions and new frontiers in rights and vision. Moreover, a prominent jurisdiction—the United States—has resisted falling in line with the GDPR's provisions, and is still searching for the right balance of protections.[85]

B. Consumer Data Privacy in the Wild West (United States)

Critics sometimes call the United States a "wild west" frontier when it comes to personal data and privacy protection.[86] In contrast to the concerted supranational coordination in the European Union, consumer data privacy protection in the United States is still a largely

---

[81] Dove, *supra* note 79, at 1013.

[82] Phillips & Knoppers, *supra* note 45, at 108.

[83] *See* ELIZABETH HANSEN SHAPIRO, MICHAEL SUGARMAN, FERNANDO BERMEJO & ETHAN ZUCKERMAN, NETGAIN PARTNERSHIP, NEW APPROACHES TO PLATFORM DATA RESEARCH 28–39 (2021), https://drive.google.com/file/d/1bPsMbaBXAROUYVesaN3dCtfaZpXZgI0x/view (last visited Sept. 13, 2021) (discussing current strategies for data collection both with and without platform cooperation and the problems associated with these methods).

[84] *See* Clare Sullivan, *GDPR Regulation of AI and Deep Learning in the Context of IOT Data Processing—A Risky Strategy*, J. INTERNET L., Dec. 2018, at 1, 19.

[85] *See infra* Section I.B; *see also* Sullivan, *supra* note 84, at 19 (identifying that the United States is a notable exception to nations falling under the influence of GDPR provisions).

[86] *See* Jennifer Huddleston, *Preserving Permissionless Innovation in Federal Data Privacy Policy*, J. INTERNET L., June 2019, at 1, 18 ("Sometimes critics allege that the United States has been a 'Wild West' when it comes to data privacy and protection.").

piecemeal patchwork.[87] Federal legislation remains in germination.[88] Depending on one's perspective, the different approach in the United States may be due to a greater focus on other important interests such as innovation—or reflective of the "weak tradition of privacy" or "weak or nonexistent privacy regime" in the United States.[89]

The nascent approach to comprehensive consumer data regulation is not for lack of interest. More than eighty percent of Americans surveyed by the Pew Research Center reported they go online daily, with most reporting they are either "almost constantly" online or online "several times a day."[90] An even larger Pew survey found that seventy-nine percent of Americans are concerned over how companies use their personal data.[91] In recent years, numerous states have considered consumer data privacy legislation.[92] This Section describes some key recent legislation and proposals that are important in spurring debate but remain largely limited in their focus on rights to one's individual data, rather than the public's right of access to big data for public interest purposes.

1. States Lead on Data Privacy Laws

With one of the most populous consumer bases in the United States, California took the lead in passing "the most far-reaching privacy measure ever to be enacted" in the nation.[93] Enacted just one

---

[87] *Id.* (noting that California's passage of a privacy protection law "potentially creat[es] a state-level patchwork").

[88] *See* discussion *infra* Section I.B.2.

[89] *Compare, e.g.*, Michael L. Rustad & Thomas H. Koenig, *Towards a Global Data Privacy Standard*, 71 FLA. L. REV. 365, 370, 411 (2019) (describing critiques of America's "weak tradition" and "weak or nonexistent privacy regime"), *with* Huddleston, *supra* note 86, at 18 (explaining the problems with European-type strict privacy regulations on innovation and preferring American openness toward innovation).

[90] Andrew Perrin & Sara Atske, *About Three-in-Ten U.S. Adults Say They Are 'Almost Constantly' Online*, PEW RSCH. CTR. (Mar. 26, 2021), https://www.pewresearch.org/fact-tank/2021/03/26/about-three-in-ten-u-s-adults-say-they-are-almost-constantly-online. Among the 1,502 Americans aged 18 or older surveyed in 2021, 79% were either "almost constantly" or "several times a day" online. *Id.*; *see also* PEW RSCH. CTR., INTERNET FREQUENCY UPDATE METHODOLOGY (2021), https://www.pewresearch.org/wp-content/uploads/2021/03/Internet-Frequency-Update-Methodology-Topline.pdf (reporting survey size and methods for the study described in this footnote).

[91] AUXIER ET AL., *supra* note 34. For information on the survey's size, see *American Trends Survey Methodology*, *supra* note 34 (reporting a total of 15,134 respondents after the 2019 and 2020 waves of recruitment).

[92] *See 2020 Consumer Data Privacy Legislation*, NAT'L CONF. OF STATE LEGISLATURES (Jan. 17, 2021), https://www.ncsl.org/research/telecommunications-and-information-technology/2020-consumer-data-privacy-legislation637290470.aspx (summarizing for each state consumer data privacy legislation that has been introduced and the outcome).

[93] Grant Davis-Denny, Jordan Navarrette & Nefi Acosta, *The California Consumer Privacy Act: 3 Early Questions*, LAW 360 (July 2, 2018, 4:28 PM), https://www.law360.com/articles/1059403/the-california-consumer-privacy-act-3-early-questions.

month after the GDPR entered into effect, the CCPA became operative in January 2020.[94] The CCPA explicitly focuses on the relationship between businesses and their consumers, whereas the GDPR is couched in terms generally applicable to all data controllers, even though its default paradigm is businesses' use of consumer data for commercial purposes.[95] The CCPA has several similarities with its wider-sweeping European counterpart in terms of the rights granted to consumers over their personal data, and the regulation of entities that may be outside the regulated jurisdiction if they also conduct business in the jurisdiction.[96]

Both the CCPA and GDPR contain bills of rights for consumers.[97] Similar to the data subject rights in the GDPR, the CCPA gives consumers rights of information, access, and control over personal data.[98] Consumers have the right to know, upon request, what personal information businesses collect about them; the sources, uses, and purposes of the information; and what third parties have received the information.[99]

In some respects, the CCPA gives consumers even greater control over their personal information. Rather than just the right of rectification of incorrect information, the CCPA gives consumers the right to request that the business delete any information the business has about the consumer.[100] The CCPA also gives consumers an "opt-out" right to direct businesses that sell personal information not to sell personal information about that consumer, on request.[101] Consumers are protected from discriminatory business practices, such as charging dif-

---

94 CAL. CIV. CODE §§ 1798.100–.199 (West 2021).

95 *See supra* text accompanying notes 81–82.

96 *See* Joanna Kessler, Note, *Data Protection in the Wake of the GDPR: California's Solution for Protecting "The World's Most Valuable Resource*," 93 S. CAL. L. REV. 99, 111–12 (2019) (describing parallels between the CCPA and the GDPR); Michael R. Overly, *Is California's Consumer Privacy Act of 2018 Going to Be GDPR Version 2?*, NAT'L L. REV., Sept. 6, 2018 (noting that the CCPA's scope is more akin to the GDPR than to traditional privacy statutes in the United States); *see also* GDPR, *supra* note 39, art. 3(2) (noting that under certain circumstances, "[t]his Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union"); CIV. § 1798.140(c)(1) (defining a regulated business as one "that does business in the State of California").

97 *Compare* GDPR, *supra* note 39, ch. III (listing rights of the data subject), *with* CIV. § 1798.150 (enumerating consumer rights and protections).

98 CIV. §§ 1798.100–.145; *see also supra* text accompanying notes 47–64 (outlining the rights in the GDPR).

99 CIV. §§ 1798.100, 1798.110, 1798.115.

100 *Id.* § 1798.105.

101 *Id.* § 1798.120.

ferent prices or offering different levels of services, based on the exercise of their consumer rights under the CCPA.[102]

The CCPA makes some provision for the use of consumer data research, defined as "scientific, systematic study and observation, including basic research or applied research that is in the public interest and that adheres to all other applicable ethics and privacy laws or studies conducted in the public interest in the area of public health."[103] The requirements for research are stringent. The CCPA requires that the data be anonymized, thoroughly protected against the possibility of intentional or inadvertent consumer reidentification, and used exclusively for noncommercial and limited research purposes.[104]

Further, the data must be "[s]ubjected *by the business conducting the research* to additional security controls that limit access to the research data to *only those individuals in a business* as are necessary to carry out the research purpose," suggesting that the CCPA contemplated regulation of for-profit businesses such as pharmaceutical companies which use consumer data.[105] Apparently omitted from the scope of the CCPA's consideration is the grant of access to consumer data by businesses to nonprofit entities for public interest research.[106]

While the CCPA notably and unusually (for an American law) emulates a more European-type model of a declaration of individual privacy rights, the CCPA is overall less comprehensive and burdensome for data gatherers than the GDPR, and its fines are less severe. The CCPA does not impose the extensive list of obligations on data controllers that is set forth in the GDPR.[107] In contrast to the potentially multimillion dollar administrative fines for some types of violations under the GDPR, the CCPA provides for:

---

[102] *Id.* § 1798.125. While consumers may not be punished for exercising their rights, businesses may offer incentives for sharing personal information, such as offering different prices or levels of service based on the value of personal information received. *Id.* § 1798.125(2)(b)(1).

[103] *Id.* § 1798.140(s).

[104] *Id.*

[105] *Id.* (emphasis added); *see also, e.g.*, *id.* § 1798.140(c) (defining business as "[a] sole proprietorship, partnership, limited liability company, corporation, association, or other legal entity that is *organized or operated for the profit or financial benefit of its shareholders or other owners* . . . and that satisfies one or more of the following thresholds" on gross revenue, or volume of personal information dealt, or proportion of annual revenue derived from selling personal information (emphasis added)).

[106] This interpretation is further suggested by the explicit exclusion of health care providers, protected health information, medical data, and information collected as part of a clinical trial from the scope of the CCPA. *See id.* § 1798.145(c)(1).

[107] *See* GDPR, *supra* note 39, ch. IV (requiring "[d]ata protection by design" through privacy impact assessments and numerous other obligations imposed on data controllers).

> [A]n injunction and liab[ility] for a civil penalty of not more than two thousand five hundred dollars ($2,500) for each violation or seven thousand five hundred dollars ($7,500) for each intentional violation, which shall be assessed and recovered in a civil action brought in the name of the people of the State of California by the Attorney General.[108]

Civil suits brought by consumers are authorized for data breaches, but damages are capped at no more than $750 per incident.[109] Also notably, unlike the GDPR, the CCPA does not tackle the thorny issue of automated decisionmaking and profiling facilitated by the amassing of big data.[110]

Among the many states to consider data privacy legislation after California's new law, Maine and Nevada succeeded in passing consumer data privacy legislation in 2019.[111] Both pieces of legislation are more modest than the CCPA. Maine's new law only applies to broadband internet providers and requires the providers to keep customer data secure and obtain explicit consumer consent before using, disclosing, selling, or providing access to the customer's personal information unless an exception applies.[112] Nevada's law requires operators of online services or internet websites to give Nevada consumers the ability to request that their information not be sold for money.[113]

Other major-market states, including Massachusetts, New York, and Washington, have considered consumer data privacy legislation in the wake of the CCPA, but thus far the proposals have not passed.[114]

---

108 CIV. § 1798.155. Fines of $2,500 (or even $750, which is the cap for civil suits brought by consumers) per violation adds up very quickly if a company's actions have affected a large number of users. *See, e.g.*, Order Granting Plaintiffs' Motion for Preliminary Approval of Class Action Settlement at 12, *In re* Hanna Andersson & Salesforce.com Data Breach Litig., No. 20-cv-00812 (N.D. Cal. Dec. 29, 2020), ECF No. 68 (preliminarily approving a $400,000 settlement in a consumer class action brought under the CCPA).

109 CIV. § 1798.150(a)(1)(A).

110 For a discussion of the GDPR's provisions on automated decisionmaking and profiling, see *supra* text accompanying notes 57–64.

111 ME. STAT. tit. 35-A, § 9301 (2021); NEV. REV. STAT. § 603A (2021); *see also* Lothar Determann & Helena J. Engfeldt, *Maine and Nevada's New Data Privacy Laws and the California Consumer Privacy Act Compared*, BAKER MCKENZIE (June 20, 2019), https://www.bakermckenzie.com/en/insight/publications/2019/06/maine-and-nevada-new-data-privacy-laws (comparing the three new state laws).

112 *See* tit. 35-A, § 9301(3)–(4) (enumerating exceptions from statutory coverage).

113 REV. § 603A.345.

114 *See, e.g.*, Khari Johnson, *Washington Privacy Act Fails Again, but State Legislature Passes Facial Recognition Regulation*, VENTUREBEAT (Mar. 12, 2020, 6:09 PM), https://venturebeat.com/2020/03/12/washington-privacy-act-fails-in-state-legislature-again (reporting the failure of proposed data privacy legislation in Washington); Peter J. Guffin, Donald R. Frederico & Melanie A. Conroy, *State Legislature Hears Concerns About Proposed Massachusetts Consumer Data Privacy Bill*, NAT'L L. REV., Oct. 11, 2019

California's recent law remains the most sweeping American effort to protect consumer data privacy.[115] While an ambitious start for U.S. consumer privacy law, the CCPA's prime focuses remain the consumer-business relationship and rights to one's personal data, rather than public access to pooled data for public interest purposes such as research by nonprofit entities.[116]

2. *Congressional Debates and Proposals*

Crafting comprehensive consumer data privacy legislation at the federal level is even more challenging in a nation with varying cultural views regarding privacy, free commerce, government regulation, protection for the vulnerable, and individual rights.[117] Amid concerns over a patchwork of state consumer data privacy laws and recurring controversies over how companies use and protect personal data, there is strong interest within Congress to address these concerns, but a lack of consensus on how to resolve key issues.[118] Congress has contemplated various proposals regarding consumer rights, regulatory burdens on businesses, remedies for violations, and other key legal

(describing substantial opposition to a proposed data privacy bill in Massachusetts); Jeewon Kim Serrato & Susan Ross, *Nevada, New York and Other States Follow California's CCPA*, DATA PROT. REP. (June 6, 2019), https://www.dataprotectionreport.com/2019/06/nevada-new-york-and-other-states-follow-californias-ccpa (describing proposed New York legislation similar to the CCPA).

115 *See supra* text accompanying notes 93–96.

116 *See supra* text accompanying notes 95–106.

117 *See, e.g.*, Sophie Cockcroft & Saphira Rekker, *The Relationship Between Culture and Information Privacy Policy*, 26 ELEC. MKTS. 55, 65–70 (2015) (analyzing cultural predictors of the level of privacy legislation in national jurisdictions); Kevin Lewis, Jason Kaufman & Nicholas Christakis, *The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network*, 14 J. COMPUT.-MEDIATED COMMC'N 79, 93–94 (2008) (discussing differing cultural preferences, of which a "taste for privacy" is only part of the influences, even among the relatively more homogenous group of U.S. students at a private college).

118 *See, e.g.*, *GDPR & CCPA: Opt-Ins, Consumer Control, and the Impact on Competition and Innovation: Hearing Before the S. Comm. on the Judiciary*, 116th Cong. (2019) (discussing bipartisan and industry interest in crafting federal data privacy legislation but splitting on approaches); Mabel Crescioni & Tara Sklar, *The Research Exemption Carve Out: Understanding Research Participants Rights Under GDPR and U.S. Data Privacy Laws*, 60 JURIMETRICS 125, 135–36 (2020) (discussing differences in approaches that render federal legislation difficult).

issues.[119] None thus far have made significant progress toward passage.[120]

Among the toughest federal contenders is the Mind Your Own Business Act of 2019, legislation introduced by Senator Ron Wyden, a Democrat from Oregon.[121] Spurred by data breach controversies, the legislation has similarly severe financial penalties like the GDPR and targets large companies as defined by gross revenues.[122] The legislation imposes numerous requirements on large businesses, including "automated decision system impact assessments" and "data protection impact assessments" and further authorizes the Federal Trade Commission to impose more regulations.[123] The bill would also expand the definition of substantial injuries from data breaches to include noneconomic injuries.[124] Commentators view the proposed legislation as unlikely to pass in its current form.[125]

Two other federal proposals take different approaches to equalizing the odd imbalance in responsibilities between commercial entities profiting from personal data and professionals such as doctors, lawyers, and researchers entrusted with personal information. The idea behind the proposed Data Care Act is to impose duties of care, loyalty, and confidentiality on online companies analogous to the duties doctors, lawyers, and bankers must exercise regarding their clients' information.[126] In addition to imposing fiduciary duties, the legislation would empower the Federal Trade Commission to interpret the scope and applicability of the duties and enforce them.[127]

A second piece of legislation, introduced in Congress by a bipartisan coalition, extends disclosure and review board requirements to "[a]ny large online operator that engages in any form of behavioral or

---

119 *See* Huddleston, *supra* note 86, at 18. *See generally* JENNIFER HUDDLESTON, MERCATUS CTR., GEORGE MASON UNIV., POLICY BRIEF: AN ANALYSIS OF RECENT FEDERAL DATA PRIVACY LEGISLATION PROPOSALS (2019), https://www.mercatus.org/system/files/huddleston_-_policy_brief_-_an_analysis_of_recent_federal_data_privacy_policy_proposals_-_v1.pdf (offering an overview of federal consumer data privacy proposals in the 115th and 116th Congresses).

120 *See* Huddleston, *supra* note 86, at 18–19.

121 Mind Your Own Business Act of 2019, S.2637, 116th Cong. (2019).

122 *Id.*

123 *Id.* §§ 5, 7.

124 *Id.* § 3.

125 Di Ai, *Sen. Wyden Introduces Federal Data Privacy Bill*, HARV. J.L. & TECH. (Oct. 30, 2019), https://jolt.law.harvard.edu/digest/sen-wyden-introduces-federal-data-privacy-bill (discussing the legislation).

126 *See* Press Release, Off. of Sen. Brian Schatz, Schatz Leads Group of 16 Senators in Reintroducing Legislation to Help Protect People's Personal Data Online (Dec. 3, 2019), https://www.schatz.senate.gov/press-releases/schatz-leads-group-of-16-senators-in-reintroducing-legislation-to-help-protect-peoples-personal-data-online.

127 *See* Data Care Act of 2019, S.2961, 116th Cong. §§ 2(d)–(e), 4 (2019).

psychological research based on the activity or data of its users."[128] This proposal addresses an odd imbalance in our present status quo: Private profit-driven entities have fewer hurdles to using consumer big data to research how to manipulate consumer behavior than nonprofit researchers generating knowledge to benefit the public.[129] Nonprofit academic researchers typically receive federal funds or are at institutions receiving federal funds, and thus are subject to regulations and ethical traditions requiring informed consent and institutional review board approval.[130] Private commercial entities seeking strategies to enhance profits using internally acquired data are not subject to these elaborate external review processes and protections.[131] Businesses conducting research on consumers also do not have to follow the public beneficence principle that typically guides institutional review board regulation of proposed research.[132]

A seemingly alluring approach to remedy this gap is to extend informed consent and institutional review boards to major commercial entities studying consumer manipulation. Adding disclosure and review board requirements has the surface appeal of equalizing the requirements between scientists and commercial entities. Notice-type laws also appeal to regulators because they seem cheaper and easier to enforce, appear to be less heavy-handed than command-and-

---

[128] Deceptive Experiences to Online Users Reduction Act, S.1084, 116th Cong. § 3(b) (2019).

[129] *See, e.g.*, 45 C.F.R. § 46.102(l) (2020) (defining research for which institutional review board review is required as "a systematic investigation . . . designed to develop or contribute to generalizable knowledge"). Studies illuminating how to reduce the incidence of disease and death contribute to generalizable knowledge; how to get you to buy more does not.

[130] *See, e.g.*, *id.* § 46.101(a) (explaining that institutional review board regulations "appl[y] to all research involving human subjects conducted, supported, or otherwise subject to regulation by any Federal department or agency that takes appropriate administrative action to make the policy applicable to such research").

[131] *See* Press Release, Off. of Sen. Mark Warner, Senators Introduce Bipartisan Legislation to Ban Manipulative 'Dark Patterns' (Apr. 9, 2019), https://www.warner.senate.gov/public/index.cfm/2019/4/senators-introduce-bipartisan-legislation-to-ban-manipulative-dark-patterns.

[132] *See, e.g.*, U.S. DEP'T OF HOMELAND SEC., THE MENLO REPORT: ETHICAL PRINCIPLES GUIDING INFORMATION AND COMMUNICATION TECHNOLOGY RESEARCH 9 (2012), https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/menlo_report_actual_formatted.pdf (explaining that the fundamental principle of beneficence undergirding institutional review board of research calls for maximizing the benefits while minimizing the harm to human subjects); NAT'L COMM'N FOR THE PROT. OF HUM. SUBJECTS OF BIOMEDICAL & BEHAV. RSCH., U.S. DEP'T OF HEALTH, EDUC., & WELFARE, THE BELMONT REPORT (1979), https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf [hereinafter THE BELMONT REPORT] (prescribing adherence to the principle of beneficence).

control regulations, and seem to legitimate what occurs after notice and consent.[133]

Yet there is a growing body of literature on the fictive utility of the disclosure-and-consent model, particularly in the electronic age where people click impatiently past the verbiage to get to the reward of the service or good sought.[134] Moreover, informed consent tackles just one facet of the challenge of regulating access to, and use of, pooled personal data and leaves larger underlying questions unresolved. There is a need to advance beyond the allure of extending informed consent and review board-type regulations and address overarching, unresolved questions about how our pooled personal data should be owned, controlled, accessed, and used.

On May 20, 2021, Massachusetts Congresswoman Lori Trahan introduced legislation that makes an early foray towards facilitating data-sharing by private companies, albeit limited in focus to targeted advertising data held by the world's largest companies.[135] The legislation is aimed at greater transparency surrounding targeted advertising on the world's most massive and powerful digital platforms, such as Google and Facebook.[136] Currently titled the U.S. Social Media Disclosure and Transparency of Advertisements Act, the proposed legislation would require platforms with more than 100 million monthly active users to share targeted advertisement data with academic researchers working with higher educational institutions.[137] While specific in its focus and aims to concerns over nontransparent and potentially problematic targeted advertising, the legislation shows the openness of lawmakers to fresh solutions that involve data

---

133 *See* M. Ryan Calo, *Against Notice Skepticism in Privacy (and Elsewhere)*, 87 NOTRE DAME L. REV. 1027, 1029 (2012).

134 *See, e.g.*, David A. Hoffman, *From Promise to Form: How Contracting Online Changes Consumers*, 91 N.Y.U. L. REV. 1595, 1597 (2016) (describing digital contracts as "increasingly the subject of satire"); Tess Wilkinson-Ryan, *A Psychological Account of Consent to Fine Print*, 99 IOWA L. REV. 1745, 1764–65 (2014) (discussing the perception that expecting consumers to read long contracts is unreasonable); Victoria C. Plaut & Robert P. Bartlett, III, *Blind Consent? A Social Psychological Investigation of Non-Readership of Click-Through Agreements*, 36 LAW & HUM. BEHAV. 293, 305–06 (2012) (noting that nonreadership of digital contracts may stem from perceptions that such contracts are all alike, do not offer any real choice, and are largely irrelevant); Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 665–729 (2011) (discussing generally the deficiencies of mandated disclosure regimes).

135 Press Release, Rep. Lori Trahan, Trahan Leads Introduction of Social Media DATA Transparency Legislation (May 20, 2021), https://trahan.house.gov/news/documentsingle.aspx?DocumentID=2112 (last visited Sept. 13, 2021).

136 *Id.*

137 Social Media Disclosure and Transparency of Advertisements Act of 2021 (Social Media DATA Act), H.R. 3451, 117th Cong. (2021), https://www.congress.gov/117/bills/hr3451/BILLS-117hr3451ih.pdf (last visited Sept. 13, 2021).

sharing. The proposal also is noteworthy in moving beyond the predominant privacy paradigm.

II
Access to Consumer Big Data: The Public's Rights in Pooled Personal Data

While many current legislative proposals focus on data privacy, there is much more to the story and to consumer concern over data collection by companies. A recent large-sample Pew Research Center survey found that 81% of Americans believe that the risks of data collection by companies outweigh the benefits.[138] Relatedly, 72% of Americans believe they personally benefit little to none from the personal data that companies gather on them.[139] Beyond privacy, an overarching concern is that while consumers bear the costs of data aggregation, they do not share in the benefits.

The concern is compounded by an imbalance in access to the valuable pooled data held by commercial entities for public-sector or nonprofit academic researchers. Absent legislative intervention, which remains nascent, access to potentially valuable data depends on whether a business wishes to grant it because of a shared interest, or perhaps a vision of corporate social responsibility.[140] Without a clear theoretical and legal framework for property rights in shared data, currently the companies that collect and control the data enjoy ownership-like rights of use and profits in the lucrative information, including the rights of potential exploitation, sale, and resale.[141]

The result is a modern form of "might makes right" mixed with the Lockean notion that the entity that appropriates a resource gets to own it.[142] As senior business executives acknowledge, "[t]hough some

---

138 Auxier et al., *supra* note 34, at 7.

139 *Id.*

140 *See, e.g.*, Birkin, *supra* note 26, at 251 (discussing private ownership of important consumer data and the three main bases for granting access).

141 *See, e.g.*, Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 Harv. L. Rev. 2055, 2069–72 (2004) (discussing the commodification and trading of personal data); Julia Alpert Gladstone, Data Mines and Battlefields: Looking at Financial Aggregators to Understand the Legal Boundaries and Ownership Rights in the Use of Personal Data, 19 J. Marshall J. Comput. & Info. L. 313, 329 (2001) (discussing how consumer profiles and databases are "a critical strategy to successful business" and "valuable intangible asset"); Jessica Littman, *Information Privacy/Information Property*, 52 Stan. L. Rev. 1283, 1284–86 (2000) (discussing the creation of digital profiles from our online behaviors and packaging of the personal information for sale).

142 There is a massive literature on John Locke's theory of appropriation. For a discussion, see, for example, Karl Widerquist, *Lockean Theories of Property: Justifications for Unilateral Appropriation*, 2 Pub. Reason 3, 5–15 (2010) (discussing interpretations of Lockean appropriation theory and critiques of it); Herman T. Tavani, *Locke, Intellectual Property Rights, and the Information Commons*, 7 Ethics & Info. Tech. 87, 88–93 (2005)

companies are open about their data practices, most prefer to keep consumers in the dark, choose control over sharing, and ask for forgiveness rather than permission."[143] The murk surrounding property rights in our personal data is a major unaddressed part of the discontent surrounding data use, control, and access. This Part tackles the property issues and makes the case for recognizing a right of the public to benefit from the big data amassed from us.

A. Property Approaches to Pooled Personal Data

In the heady early days of the Internet, prominent scholars proposed that the law vest property rights over personal data in the individual user, creating a market in which privacy rights and tradeoffs could be negotiated and crystallize.[144] The hope was that the invisible hand of the market would move via consumer choice toward the right balance of protections as consumers chose to visit businesses that advertised acceptable terms and policies.[145] The problem was that

(discussing the applicability of Lockean property theories to the information commons context); Jeremy Waldron, *Locke, Tully, and the Regulation of Property*, 32 POL. STUD. 98, 99–105 (1984) (construing Locke's argument about the natural property rights of individuals in a society).

143 Morey et al., *supra* note 37.

144 *See, e.g.*, LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE 123–34, 160–62 (1999) (arguing that cyberspace and the use of personal data can be regulated); Lawrence Lessig, *Privacy as Property*, 69 SOC. RES. 247, 261 (2002) (arguing that using traditional property rhetoric along with its associated moral rights to understand and argue for privacy would enhance privacy protections); Lawrence Lessig, *The Architecture of Privacy*, 1 VAND. J. ENT. L. & PRAC. 56, 63 (1999) ("So the trick is to construct a regime where those who would use the data internalize this [information] cost, by paying those whose data are used."); Pamela Samuelson, *A New Kind of Privacy? Regulating Uses of Personal Data in the Global Information Economy*, 87 CALIF. L. REV. 751, 769–73 (1999) (discussing the feasibility of establishing an individual property right to personal data); Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193, 1246–49 (1998) (analyzing the efficacy of a market solution to information privacy invasions); Kenneth C. Laudon, *Extensions to the Theory of Market and Privacy: Mechanics of Pricing Information*, NAT'L TELECOMMS. & INFO. ADMIN., U.S. DEP'T OF COM. (June 12, 1997), https://www.ntia.doc.gov/page/chapter-1-theory-markets-and-privacy (arguing that the lack of individual property rights over personal data has led to a personal information market failure).

145 *See, e.g.*, Peter P. Swire, *Markets, Self-Regulation, and Government Enforcement in the Protection of Personal Information*, NAT'L TELECOMMS. & INFO. ADMIN., U.S. DEP'T COM. (June 12, 1997), https://www.ntia.doc.gov/page/chapter-1-theory-markets-and-privacy (discussing how consumer preferences influence privacy policies under a market self-regulation model); Steven A. Bibas, Writing Competition Winner, *A Contractual Approach to Data Privacy*, 17 HARV. J.L. & PUB. POL'Y 591, 605 (1994) ("In the case of regulation of the information industry, perceptions and valuations of the privacy problem vary too greatly for a conventional, centralized solution to fit well. We must therefore turn to the branch of the common law most sensitive to individual preferences: contracts."). The idea of creating property rights in one's personal information as a strategy to protect against harms from appropriation of our information predates the big data era. *See, e.g.*, ALAN F.

people soon lost control over their personal data from their online browsing and took to clicking past shrink-wrap notices.[146] Experience also shows people are not sophisticated comparison shoppers among privacy policies.[147]

Individual privacy-oriented theories of property rights in personal data also do not address property rights in the most valuable, powerful commodity today—our pooled data. The personal data droplets we leave in our everyday Internet traverses accumulate into valuable pools of information at the individual and group level.[148] The volume and value of personal data are growing exponentially with the advance of the Internet of Things and alluring devices that collect not just identities, financial information, and contact information but also biometric and health data such as facial and fingerprint scans, sleep patterns, fitness indicators, biochemical blood data, disease markers, and more.[149]

1. Property by Capture: Intellectual Property Protections for Compiled Data

In the absence of clear laws regarding property rights in our personal data, the bundles of rights associated with ownership in our consumer big data are held by the companies that collect and control the information.[150] Consumer data are typically lucrative, privately controlled assets to be used and potentially sold for profit by the compa-

---

WESTIN, PRIVACY AND FREEDOM 323 (1968) ("[P]ersonal information, thought of as the right of decision over one's private personality, should be defined as a property right, with all the restraints on interference by public or private authorities and due-process guarantees that our law of property has been so skillful in devising."). *But see* ARTHUR R. MILLER, THE ASSAULT ON PRIVACY: COMPUTERS, DATA BANKS, AND DOSSIERS 211–12 (1971) ("[R]eal and personal property concepts are irrelevant to the personal values that we are attempting to preserve by recognizing a right of privacy.").

146 *See, e.g.*, Neil Weinstock Netanel, *Cyberspace Self-Governance: A Skeptical View from Liberal Democratic Theory*, 88 CALIF. L. REV. 395, 476 (2000) ("We are used to relinquishing control over bits of personal information in many seemingly unrelated contexts.").

147 *See, e.g.*, Littman, *supra* note 141, at 1287 ("Self-regulation is an abject failure . . . ."); Netanel, *supra* note 146, at 476 ("Internet users awash in an overabundance of information are no more able to assess and compare products and rule regimes than are their offline counterparts.").

148 *See, e.g.*, DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 13–22 (2004) (discussing the accumulation of our personal information via the digital trails we leave); Netanel, *supra* note 146, at 476–77 (noting that accumulation of many "innocuous isolated instances of data collection, spread out over a considerable period," can be "aggregated and compiled into a highly pervasive profile").

149 Elvy, *supra* note 5, at 426–27.

150 *See, e.g.*, Schwartz, *supra* note 141, at 2056–57 (explaining that companies view personal data "as a corporate asset and have invested heavily in software that facilitates the collection of consumer information").

nies that provide the enticement and infrastructure for our online interactions.[151] Professor Edward Janger argues that the murky state of property rights and privacy protections in our personal information has led to a tragedy of the commons in which data are available for the taking, leading to overuse and misuse harms.[152] Yet the state of access and control to our pooled personal data today is hardly a commons available to all. Property rights to our pooled personal data have arisen through capture and possession—a "natural expedient" default allocation of entitlements, in the words of philosopher David Hume, writing in the 1700s.[153]

Applying principles of copyright and trade secret law, courts have recognized property rights in the consumer information that companies compile.[154] The Constitution authorizes Congress to "secur[e] for limited Times to Authors . . . the exclusive Right to their respective Writings"—what we term copyright protection today.[155] Under the Copyright Act, "original works of authorship" are accorded intellectual property protection.[156] To be protectible as "original," a work must be "independently created by the author" and have "at least some minimal degree of creativity."[157] Facts such as someone's name, address, and contact information are not copyrightable.[158] A compilation or collection of facts, however, can be deemed sufficiently orig-

---

151 Nguyen, *supra* note 5, at 564–67; *see also, e.g.*, Gladstone, *supra* note 141, at 329.

152 Edward J. Janger, *Privacy Property, Information Costs, and the Anticommons*, 54 HASTINGS L.J. 899, 911 (2003).

153 *See* DAVID HUME, A TREATISE OF HUMAN NATURE 503 (Lewis Amherst Selby-Bigge & P.H. Nidditch eds., 2d ed. 1978) (observing that an inevitable problem a developing society will face is how to separate possessions: "This difficulty will not detain them long; but it must immediately occur, as the most natural expedient, that everyone continues to enjoy what he is at present master of, and that property or constant possession be conjoin'd to the immediate possession"); *see also* ROBERT SUGDEN, THE ECONOMICS OF RIGHTS, COOPERATION AND WELFARE 93–99 (2004) (noting the theoretical and philosophical roots of recognition of property rights by possession).

154 *See, e.g.*, Experian Info. Sols., Inc. v. Nationwide Mktg. Servs., Inc., 893 F.3d 1176, 1179–80, 1186 (9th Cir. 2018) (recognizing copyright protections for Experian's compilation of consumer data); *In re* Nw. Airlines Priv. Litig., No. Civ. 04-126, 2004 WL 1278459, at *4 (D. Minn. June 6, 2004) (holding that the compilation of passenger data into a record by Northwest Airlines is the company's property); Mason v. Montgomery Data, Inc., 967 F.2d 135, 136, 140–41 (5th Cir. 1992) (holding that a compilation of real estate ownership data superimposed on maps was protectible because the compiler made independent choices "to select information from numerous and sometimes conflicting sources" such as various public records and combined the data onto "an effective pictorial expression").

155 U.S. CONST. art. I, § 8, cl. 8.

156 17 U.S.C. § 102(c).

157 Feist Publ'ns, Inc. v. Rural Tel. Serv., 499 U.S. 340, 345 (1991) (establishing that information alone, without a degree of original creativity, is not protectable by copyright).

158 *Id.* at 361.

inal to receive copyright protection.[159] What is protected is the contribution by the author in selecting, compiling, or arranging the data—and that contribution must be more creative than just arranging names and phone numbers of all people in a service area alphabetically.[160]

For example, the Ninth Circuit extended copyright protections to a database of consumer personal information compiled by Experian and sold to commercial clients wanting to expand their marketing reach.[161] Experian's ConsumerView database had more than 250 million consumer records containing hundreds of datapoints about each consumer, such as purchasing habits, earnings, and behavior predictions, compiled from 2,200 public and proprietary sources of information.[162] The Ninth Circuit noted that Experian tested the data quality of each source of information and selected name and address pairings that the company believed would be valuable to clients wishing to purchase the compiled personal data, excluding the information of prisoners or the very elderly.[163] The company also deployed thousands of "business rules" or algorithms to resolve conflicts between sources and determine the information to include in the database.[164] The Ninth Circuit ruled that the company's culling of data from multiple sources, sorting through conflicts, and exercising judgment about what data to include and exclude constituted independent choices that "more than m[et] that standard" of creativity for copyright protection.[165]

Another source of property-like protections for collections of consumer personal data is trade secret law—if the company protects against public disclosure.[166] Information kept secret that has independent economic value is accorded trade secret protection.[167] Information can be considered a trade secret if: (1) it confers competitive

---

159 *See* 17 U.S.C. § 103; *see also Feist*, 499 U.S. at 347–48 (explaining that facts are "part of the public domain, available to every person" but "[f]actual compilations, on the other hand, may possess the requisite originality [to be copyrightable]" (internal quotation marks omitted) (citing Miller v. Universal City Studios, Inc., 650 F.2d 1365, 1369–70 (5th Cir. July 1981))).

160 *See Feist*, 499 U.S. at 362–63; 17 U.S.C. § 103(b).

161 Experian Info. Sols., Inc. v. Nationwide Mktg. Servs., Inc., 893 F.3d 1176, 1179–80, 1186 (9th Cir. 2008).

162 *Id.* at 1180.

163 *Id.*

164 *Id.*

165 *Id.* at 1185.

166 *See* Ruckelshaus v. Monsanto Co., 467 U.S. 986, 1011 (1984) (explaining that information must be kept secret to receive protection).

167 *See* 1 ROGER M. MILGRIM & ERIC E. BENSEN, MILGRIM ON TRADE SECRETS §§ 1.03, 1.07A (2021).

advantage or has value derived from exclusive possession, (2) a company takes reasonable means to keep it secret, and (3) it is not publicly available.[168] Traditional examples of protected trade secrets are the formula for Coca Cola Classic or the recipe of the eleven herbs and spices that flavor Kentucky Fried Chicken.[169] Experian's ConsumerView database of personal information also offers an example of trade secret protection in the context of commercializing consumer personal information. The Ninth Circuit held that Experian made a prima facie showing that its ConsumerView database also was protected under trade secret law because the company maintained the secrecy of the database.[170] While Experian sold access to the database to other entities, the access was granted under "strict security agreements with licensees to maintain the database's secrecy."[171] Thus, trade secret law rewards the efforts of companies to keep tight control over the personal data collected through property protections.

2. *Beyond Ownership by Capture and Personal Data as* Ferae Naturae

Currently, property law does not recognize individual ownership rights in one's personal data; rather property rights attach when labor is mixed with the data in a collection effort.[172] As Professor Vera Bergelson has observed, personal data in America is treated akin to wild animals or other unowned resources free for the bagging, reminiscent of the famous case of *Pierson v. Post*.[173] *Pierson* famously expounded that wild animals, termed *ferae naturae*, such as the fox pursued by hunters in the case, are up for grabs until someone takes possession by mortally wounding or seizing them.[174] In our modern times, the personal data that teem in the wilds of our online traverses are akin to the *ferae naturae* of old, lacking property protections until a company collects and aggregates them into a valuable commodity.

Proposing a hybrid model of propertized personal data, Professor Paul Schwartz noted that property is "a bundle of interests rather than

---

168 David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 145 (2007) (arguing against trade secret protections for public infrastructure).

169 David R. Hannah, *Keeping Trade Secrets Secret*, 47 MIT SLOAN MGMT. REV. 17, 17 (2006).

170 *Experian*, 894 F.3d at 1188.

171 *Id.*

172 Vera Bergelson, *It's Personal but Is It Mine? Toward Property Rights in Personal Information*, 37 U.C. DAVIS L. REV. 379, 403 (2003) ("Currently, neither property nor torts theory recognizes individuals' rights in their information.").

173 *Id.* at 403; Pierson v. Post, 3 Cai. 175 (N.Y. Sup. Ct. 1805).

174 *Pierson*, 3 Cai. at 175.

despotic dominion over a thing."[175] Under Schwartz's proposal, people would have the right to transfer their data to businesses, but further transfers such as resale of data are forbidden absent an opt-in from the individual.[176] Schwartz's proposal has such enduring practical and theoretical appeal that nearly a decade later, a former Facebook public policy director turned senior legislative counsel for the ACLU endorsed the idea as a vision for where the law should go.[177]

The current conventions on ownership of our data have arisen because of the lack of clear law regarding property rights in our personal data. As a result, some of the most cutting-edge questions of our times are relegated to remarkably primitive expedient defaults such as possession being nine-tenths of the law.[178] The property-like protections in our data that courts do recognize are squeezed into legal doctrines fashioned for simpler kinds of property with less far-reaching public interest as well as private ramifications. It does not have to be this way. The fomenting debate trying to address public concerns over how personal data are controlled and used presents an opportunity to also address root problems in the distribution of rights to benefit from our pooled personal information.

B. Toward a Public Right to Benefit from Our Pooled Data

The current debates around property protections for our personal data has a gap between frames. The debate about property protections for personal data, usually spurred by concerns over privacy harms, focuses on an individual's ownership of personal data points—just drops in the vast ocean of big data. Judicially accorded property protections for compilations of data focus on the data collector's rights, not the rights of the public whose data is compiled into a valuable commodity. Missing is consideration of the public's rights in our pooled data.

The odd state of affairs is illustrated by application to the metaphor of the Stone Soup at the outset of the Article.[179] Under the current approach to our pooled data, the exclusive control to enjoy the

---

175 Schwartz, *supra* note 141, at 2094.

176 *Id.* at 2095–106.

177 *See* Timothy D. Sparapani, *Putting Consumers at the Heart of the Social Media Revolution: Toward a Personal Property Interest to Protect Privacy*, 90 N.C. L. REV. 1309, 1313 (2012) ("[Schwartz] advocated for what he termed a 'use-transfer restriction' regarding personal data, and I endorse that limitation.").

178 For a discussion of the intellectual origins of the old adage, see Carol M. Rose, *The Law Is Nine-Tenths of Possession: An Adage Turned on Its Head*, *in* LAW AND ECONOMICS OF POSSESSION 40 (Yun-Chien Chang ed., 2015).

179 *See supra* notes 1–4 and accompanying text.

Stone Soup would be held by the travelers who provided the pot and enticed the villagers out with their contributions. Moreover, the travelers may not share the soup with the villagers whose contributions comprise the soup if they want to keep their rights. The incentive to maintain exclusive control of powerful pooled data for revenue-enhancing purposes is intensified by trade secret law, which only protects information that a company succeeds in keeping secret and out of the public domain.[180]

Recognizing a right of the public to benefit does not mean releasing a company's valuable commodity into the public domain. Nor does it mean vitiating privacy protections and data security safeguards. Rather, it means carving out provisions for public-interest access and safe harbors that facilitate data sharing with parties qualified and trained to protect the data and carry out research that creates public benefits such as improving health or safety. Such sharing is in the interest of businesses as well as the community to improve public perceptions and build trust regarding how personal data are being used.[181] It also represents a more just and equitable allocation of property interests in our pooled personal data—and the power that access to the information confers.[182]

1. Access for Public Benefit Purposes

The consumer data that businesses use to target advertising, expand sales, and encourage spending also could be productively deployed for public benefit, to prevent outbreaks of disease, trace and address the spread of false and dangerous claims, and more.[183] For

---

[180] *See, e.g.*, Ruckelshaus v. Monsanto Co., 467 U.S. 986, 1011 (1984) ("Once the data that constitute a trade secret are disclosed to others, or others are allowed to use those data, the holder of the trade secret has lost his property interest in the data.").

[181] *See, e.g.*, Morey et al., *supra* note 37 (arguing that gaining customers' confidence will be key in the developing consumer-data space, as "[c]ompanies that are transparent about the information they gather, give customers control of their personal data, and offer fair value in return for it will be trusted and will earn ongoing and even expanded access").

[182] *See, e.g.*, Alessandro Mantelero, *Social Control, Transparency, and Participation in the Big Data World*, 17 J. INTERNET L. 23, 25–26 (2014) (discussing the need to "limit the power over information and obtain a better allocation of it," "in order to limit possible abuse and illegitimate advantages [and] . . . . increase access to information[,] . . . spreading the informational power currently in the hands of a few bodies").

[183] *See, e.g.*, Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant, *Detecting Influenza Epidemics Using Search Engine Query Data*, 457 NATURE 1012, 1012–14 (2009) (presenting a method to track the spread of influenza-like illness using high volumes of Google search query data); *Anti-Racism Protests: Divisive Disinformation Narratives Go Viral on Facebook, Racking Up Over 26 Million Estimated Views*, AVAAZ (June 12, 2020), https://secure.avaaz.org/campaign/en/anti_protest_disinformation (reporting on the spread of misinformation about protests over race and policing using Facebook data).

example, the granular user location data that Google deploys for targeted advertising, product development, and other commercial strategies also can be used to evaluate the success of public health campaigns to encourage social distancing and limit travel to curb the COVID-19 pandemic.[184] Amassed cell phone location data also can be used to evaluate compliance with COVID-19 reduction measures restricting travel and social gatherings.[185] Google search query data have also created new ways to track the spread of influenza and detect epidemics.[186] Data released by Facebook have been used by a watchdog group to study the spread of false information about vaccines and other health facts and to identify some of the major sources of the spread of misinformation imperiling health.[187] With the proliferation of more online services and devices in the Internet of Things, the volume of data and possibilities for public interest research also will continue to exponentially expand.[188]

Currently, access for external researchers to privately held consumer data is a matter of voluntary grace by private companies—motivated by anticipated benefits to the company, philanthropy, and perhaps a sense of social responsibility.[189] While some companies have become more progressive about voluntarily sharing data, valuable consumer big data too often remains locked away for private profitmaking, its public interest potential unrealized.[190] Attempts by researchers to access privately held big datasets and their vast poten-

---

184 Natasha Lomas, *Google Is Now Publishing Coronavirus Mobility Reports, Feeding Off Users' Location History*, TechCrunch (Apr. 3, 2020, 9:19 AM), https://techcrunch.com/2020/04/03/google-is-now-publishing-coronavirus-mobility-reports-feeding-off-users-location-history.

185 *See* Nigel Chiwaya, *Analysis: Data from 15 Million Phones Shows Some Americans Are Gathering at Pre-Pandemic Levels*, NBC News (June 11, 2020, 6:31 AM), https://www.nbcnews.com/news/us-news/analysis-data-15m-phones-shows-some-americans-are-gathering-pre-n1229636.

186 *See* Ginsberg et al., *supra* note 183, at 1012–14.

187 *See Facebook's Algorithm: A Major Threat to Public Health*, AVAAZ (Aug. 19, 2020), https://avaazimages.avaaz.org/facebook_threat_health.pdf; *see also* Elizabeth Dwoskin, *Misinformation About the Coronavirus Is Thwarting Facebook's Best Efforts to Catch It*, Wash. Post (Aug. 19, 2020, 6:00 AM) (reporting on the "left-leaning" group AVAAZ's study using Facebook's released data about the spread of health misinformation during the COVID-19 pandemic).

188 *See, e.g.*, Sylvia Zhang, *Who Owns the Data Generated by Your Smart Car?*, 32 Harv. J.L. & Tech. 299, 299–300 (2018) (discussing new opportunities to study ways to improve traffic safety and our public infrastructure using autonomous vehicle data).

189 *See, e.g.*, Leslie Harris & Chinmayi Sharma, Future of Priv. F., Understanding Corporate Data Sharing Decisions 7–8 (2017), https://fpf.org/wp-content/uploads/2017/11/FPF_Data_Sharing_Report_FINAL.pdf (examining the reasons why companies share data with researchers); Birkin, *supra* note 26, at 251 (summarizing the primary reasons why companies share data with researchers).

190 *See* Birkin, *supra* note 26, at 251.

tial often founder on proprietary barriers, trade secret concerns, and the shoals of consumer privacy laws, which, as discussed in Part I, are usually framed with business uses of the data as the paradigmatic concern.[191]

As a Future of Privacy Forum study reports, "[m]ost corporate data is typically unavailable to academic researchers."[192] Numerous academic researchers interviewed in the exploratory study "expressed concern about the unavailability of corporate data in any form."[193] Moreover, as companies accelerate in amassing valuable data, "researchers also expressed concern that companies rather than social scientists will increasingly set the research agenda."[194]

Relegating access to our pooled data to the whims and willingness of business is problematic for several reasons. First, there is a risk of selection bias in both the types of studies for which access is granted—and the researchers who receive access.[195] Access may only be granted to researchers who have views or hypotheses agreeable to the company. Indeed, access to data may even be a recruiting tool for business to attract talent to affiliate with the company.[196] Similarly, only certain sets of data may be made available, potentially dictating the research agenda for what can be studied—and even potentially biasing the outcomes of results.

Even after data are provided, studies that yield embarrassing results for a business may lead to restrictions on access or loss of access to the data. Even the risk of lost access may lead researchers to be cautious about publishing results that might risk the ire of the gatekeepers to valuable data. Even worse, access may be conditioned on reporting only certain kinds of favorable findings or prepublication review. Overarching all these concerns, even if there is no conflict or bias, there remains a risk of the *perception* of conflict or bias that harms trust in the research and researchers.

Even when shared, data may be too limited in format, content, or permissions to enable the study of certain important questions such as

---

191 *See* Gary King & Nathaniel Persily, *A New Model for Industry-Academic Partnerships*, 53 PS: POL. SCI. & POL. 703 (2020) [hereinafter King & Persily, *A New Model*].

192 HARRIS & SHARMA, *supra* note 189, at 1.

193 *Id.* at 4.

194 *Id.*

195 *See id.* at 4 (discussing lengthy trust-building required to woo companies into sharing data).

196 *See, e.g.*, Matt Stempeck, *Sharing Data Is a Form of Corporate Philanthropy*, HARV. BUS. REV. (July 24, 2014), https://hbr.org/2014/07/sharing-data-is-a-form-of-corporate-philanthropy (noting the "earned media opportunities, free labor represented by academics' brains, and the potential to hire the valuable talent that can emerge from such partnerships").

the impact of poverty on important outcomes.[197] Access to researchers also may become a commodity for which prohibitive prices may be charged, potentially shutting out public interest research that is not well-funded.[198] Some business might choose not to share data at all, meaning that the public is locked out of the benefits of their pooled data. Companies that are more progressive in sharing data might also risk backlash for doing so.[199] A recognized legal right to access for public interest research may actually take heat off companies from consumers for granting access to valuable information for the benefit of the public.

2. *Deciding What Constitutes a Public Interest Use*

A statutory solution would be to explicitly recognize a right of access to pooled consumer data held by private entities for research in the public interest. The statutory duty would apply to companies that amass large volumes of consumer information during delivery of goods and services, with exceptions for small businesses that cannot bear the expenses of the additional regulatory burdens of facilitating data access, review of proposals, and sharing.

Abstract, often interchangeably used terms such as "public benefit," "public interest," "common good," and "public good," are often used in public policy and law despite definitional and procedural challenges in determining what will benefit the public overall.[200] Defining the public interest is particularly a challenge in partisan times when one person's public interest is another person's political agenda-pushing.[201] The rapid spread of misinformation means that even scien-

---

197 *See* HARRIS & SHARMA, *supra* note 189, at 4 (noting researcher concerns about the limitations of data that are shared and the potential impact on poverty research).

198 *See, e.g.*, Stempeck, *supra* note 196 ("Twitter . . . sells access to a range of real-time APIs to marketing platforms, but the price point often exceeds researchers' budgets. To accelerate the pursuit of knowledge, Twitter has piloted a program called Data Grants offering access to segments of their real-time global trove to select groups of researchers.").

199 Sarah Zhang, *Big Pharma Would Like Your DNA*, ATLANTIC (July 27, 2018), https://www.theatlantic.com/science/archive/2018/07/big-pharma-dna/566240 (discussing the backlash to 23andMe, a DNA testing company, as a result of sharing consumer genetic information with drug-developing companies).

200 *See, e.g.*, INST. OF CHARTERED ACCTS. IN ENG. & WALES (ICAEW), ACTING IN THE PUBLIC INTEREST: A FRAMEWORK FOR ANALYSIS 2–12 (2012), https://www.icaew.com/-/media/corporate/files/technical/ethics/public-int-rep-web.ashx?la=en [hereinafter ICAEW, ACTING IN THE PUBLIC INTEREST] (offering a deep discussion of such challenges for a professional body of accountants, showing that these definitional and procedural challenges exist across professional domains).

201 *See, e.g.*, *id.* at 2 ("Invoking the public interest requires justification of an ability and right to decide what is for the greater good, in the face of a natural suspicion that those proposing an action in the public interest are actually acting in their own interests."); Jane

tifically backed public health measures, such as those aimed at curbing a pandemic, are challenged as political ploys or conspiracies to undermine rather than protect the public.[202] Ideologically motivated reasoning leads people to discount policy purposes that do not conform to preexisting beliefs and amplify other, potentially spurious claims or causes that conform with prior beliefs.[203] The inevitable risk of contestation over what counts as public interest is an important consideration—but one that the law navigates all the time in diverse domains and even different legal jurisdictions.[204]

Johnston, *Whose Interests? Why Defining the 'Public Interest' Is Such a Challenge*, CONVERSATION, https://theconversation.com/whose-interests-why-defining-the-public-interest-is-such-a-challenge-84278 (Jan. 22, 2019, 7:17 PM) (explaining how oft-used terms such as "public good," "public interest," "common interest," and "common good" are hard to precisely define—a task that has sparked "[c]enturies of scholarship" from "some big names in political philosophy").

202 *See, e.g.*, Jessica Jaiswal, Caleb LoSchiavo & David C. Perlman, *Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism*, 24 AIDS & BEHAV. 2776, 2777 (May 21, 2020) (discussing resistance to COVID-19 public health measures and the spread of misinformation (potentially inadvertent falsehoods) and even deliberate disinformation); Fabio Tagliabue, Luca Galassi & Pierpaolo Mariani, *The "Pandemic" of Disinformation in COVID-19*, 2 SN COMPREHENSIVE CLINICAL MED. 1287, 1287–88 (2020) (discussing the spread of "conspiracy or denial ideas" regarding COVID-19); WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse & IFRC, Joint Statement: Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation (Sept. 23, 2020), https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation (discussing how misinformation and disinformation on public health responses "costs lives" by "amplifying hate speech; heightening the risk of conflict, violence and human rights violations; and threatening long-terms [sic] prospects for advancing democracy, human rights and social cohesion").

203 *See, e.g.*, Toby Bolsen, James N. Druckman & Fay Lomax Cook, *The Influence of Partisan Motivated Reasoning on Public Opinion*, 36 POL. BEHAV. 235, 237 (2014) (explaining how "[m]otivated directional reasoning causes people to seek out information that confirms their existing beliefs . . . , counter-argue and dismiss information inconsistent with their existing beliefs regardless of the belief's objective accuracy . . . , and view evidence consistent with their prior opinions as stronger"); Erik C. Nisbet, Kathryn E. Cooper & R. Kelly Garrett, *The Partisan Brain: How Dissonant Science Messages Lead Conservatives and Liberals to (Dis)Trust Science*, 658 ANNALS AM. ACAD. POL. & SOC. SCI. 36, 37–40 (2015) (explaining the impact of ideologically motivated reasoning in partisan perceptions of whether policies are beneficial or a waste of money).

204 *See, e.g.*, Australian Law Reform Comm'n, *Serious Invasions of Privacy in the Digital Era* (Discussion Paper No. 80, Mar. 2014) § 31.03.2014, proposal 8–2 (proposing a list of public interest considerations such as public health, national security, economic wellbeing, and freedom of expression that Australian courts should consider in enforcing privacy protections); *Media: Assessing the Public Interest in Cases Affecting the Media*, CPS, https://www.cps.gov.uk/legal-guidance/media-assessing-public-interest-cases-affecting-media (Nov. 11, 2019) (defining factors that UK prosecutors must consider to determine whether a prosecution of media actors is in the public interest); Stephen M. King, Bradley S. Chilton & Gary E. Roberts, *Reflections on Defining the Public Interest*, 41 ADMIN. & SOC'Y 954, 955 (2009) (outlining how the Federal Communications Commission applies its "public interest" standard).

Administrative agencies such as the historical Interstate Commerce Commission, the Federal Trade Commission, and the Federal Communications Commission (FCC), have long used the "public interest" as a standard to guide action.[205] As the administrative state grew after the Industrial Revolution and during World War II, the idea of the public interest was "a central concern in the exercise of discretion and the responsible use of expertise in determining and modifying public policies."[206] Today, agencies such as the FCC continue to implement laws and regulations that use the public interest as a standard.[207] For example, the Communications Act of 1934, as amended, requires that the FCC allocate licenses to use the radio and television broadcast spectrum based on a finding "that public interest, convenience, and necessity would be served by the granting thereof."[208] The public interest standard "no doubt leaves wide discretion and calls for imaginative interpretation."[209] The important term "public interest" was not defined by Congress; rather the FCC was entrusted with developing the meaning in its decision-making processes.[210]

Despite a venerable history of scholars wrestling with the term "public interest" and its frequent synonyms such as "public benefit," these concepts remain firmly ensconced in law and policy because, as political scientist Richard E. Flathman put it, "[w]e are free to abandon the *concept*, but if we do so we will simply have to wrestle with the problems under some other heading."[211]

Rather than fixed by one definition or frozen in time, the concept of what constitutes the public interest is under constant review and

---

[205] *See, e.g.*, PENDLETON HERRING, PUBLIC ADMINISTRATION AND THE PUBLIC INTEREST 138–65 (1967 ed. 1936) (discussing the widespread use of the concept among agencies as early as 1936 despite the dearth of a definition); Orion F. White & Cynthia J. McSwain, *The Phoenix Project: Raising a New Image of Public Administration from the Ashes of the Past*, 22 ADMIN. & SOC'Y 3, 4–29 (1990) (discussing the use of the concept of the public interest to guide administrative action).

[206] Gary S. Marshall & Enamul Choudhury, *Public Administration and the Public Interest: Re-Presenting a Lost Concept*, 41 AM. BEHAV. SCIENTIST 119, 122 (1997).

[207] King et al., *supra* note 204, at 955 (investigating "the public interest standard of the Federal Communications Commission").

[208] Communications Act of 1934, §§ 309(a), 310(d), 48 Stat. 1064 (codified as amended at 47 U.S.C. § 309(a)).

[209] FCC v. RCA Commc'ns, Inc., 346 U.S. 86, 90 (1953).

[210] Communications Act of 1934 § 303(r) (codified as amended at 47 U.S.C. § 303(r)) (providing that "the Commission from time to time, as public convenience, interest, or necessity requires, shall . . . [m]ake such rules and regulations and prescribe such restrictions and conditions, not inconsistent with law, as may be necessary to carry out the provisions of [the Act]").

[211] RICHARD E. FLATHMAN, THE PUBLIC INTEREST: AN ESSAY CONCERNING THE NORMATIVE DISCOURSE OF POLITICS 13 (1966).

determination through application.[212] Indeed, some experts have counseled that it is better not to define the term public interest or public benefit, instead providing a nonexhaustive list of factors to consider such as public health, safety, and economic wellbeing to guide interpretation via adjudication.[213]

The European Commission's proposal to streamline consent and access to data released for "altruistic purposes" offers insights into a broad open-ended approach to defining what constitutes a public good purpose.[214] The proposed Data Governance Act refers to consensual data releases for "purposes of general interest such as scientific research purposes or improving public services."[215]

Ultimately, determining what constitutes a public interest use is a procedural design issue, rather than a definitional one because procedural fairness confers legitimacy in policy and decisionmaking.[216] Two crucial procedural design issues are: (1) who or what gets to decide whether a proposed use is in the public interest, and (2) under what procedure? Fortunately, these important design questions do not face a blank slate—rather, there are venerable, time-tested procedures for review of research proposals that weigh the anticipated public benefit from research against the potential harms of accessing potentially sensitive data. The next Part details the institutional design and review lessons from institutional review board scrutiny of research proposals seeking to access and use data. These procedures that underpin advances in science and medicine can inform the design of procedures for determining who gets to access and use pooled consumer data for public benefit purposes and how to redress the risk of privacy and related harms.

---

[212] GARY L. WAMSLEY, ROBERT N. BACHER, CHARLES T. GOODSELL, PHILIP S. KRONENBERG, JOHN A. ROHR, CAMILLA M. STIVERS, ORION F. WHITE & JAMES F. WOLF, REFOUNDING PUBLIC ADMINISTRATION 40–41 (1990).

[213] Australian Law Reform Comm'n, *supra* note 204, § 31.03.2014, prop. 8-2.

[214] *Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*, COM (2020), 767 final, 15–22 (Nov. 25, 2020).

[215] *Id.* at 23.

[216] *See, e.g.*, Tom R. Tyler, *Governing amid Diversity: The Effect of Fair Decision-Making Procedures on the Legitimacy of Government*, 28 LAW & SOC'Y REV. 809, 810–12 (1994) (discussing the impact of procedural justice on perceptions of legitimacy of policies and outcomes).

III
Realizing a Public Right to Benefit from Privately Held Consumer Big Data

Recognizing a right of public interest access to our common resource of pooled personal information presents important questions of institutional and procedural design to maximize benefits while reducing harms. Sharing consumer data for common-benefit purposes presents the risk of a big data era version of the tragedy of the commons.[217] This is not insurmountable. Rather, we can draw insights from medical and health science research, which offer time-tested procedures for reviewing claims of public benefit from using sensitive data and offering access while safeguarding some of our most sensitive information.[218] Strategies of controlled access—ensuring that only those qualified and competent to use and protect data may obtain the information—can provide the benefits of big data sharing while preventing privacy and related harms.[219] This Part proposes a controlled-access model of data sharing and also recommends creating statutory safe harbors or regulatory sandboxes so that businesses have the right incentives to share useable data to enrich the public, not just retain key information for private profit.

A. Preventing the Tragedy of the Data Commons: Privacy and Related Harms

Unlike tangible natural resources like fisheries or rangelands, data do not get used up when shared—constituting what economics and property scholars call a "nonrivalrous" good.[220] Because data do not diminish when shared, in some ways they seem particularly well-suited for the freedom of a commons, avoiding the classic tragedy of the commons in which unconstrained use leads to depletion and destruction of the resource.[221] Historically, commons treatment was

---

217 *Cf.* Garrett Hardin, *The Tragedy of the Commons*, 162 Science 1243, 1244 (1968) (offering the classical account of the tragedy of freedom in a commons with the example of cattle overgrazing rangeland because of uncontrolled use).

218 *See* discussion *infra* Section III.B.

219 *See, e.g.*, Fan, Camera Power, *supra* note 3, at 195–204 (discussing how controlled access can facilitate access to police-worn body camera videos for violence prevention and civil liberties protection research while protecting against privacy harms).

220 *See, e.g.*, Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 Ariz. L. Rev. 339, 373 (2017) (explaining that data are "nonrivalrous, because [they] can easily and cheaply be copied and shared, at least technologically" and "data collectors and analyzers have the potential to sell or license their data sets to multiple users").

221 *See, e.g.*, James M. Acheson, The Lobster Gangs of Maine 48–49, 73–76, 142–44 (1988) (discussing the problems with open-access and intermediate communal property regimes using the example of the lobster gangs of Maine); Hardin, *supra* note 217, at 1244 (discussing resource depletion as the tragedy of freedom in a commons).

accorded to "plenteous goods"—goods like the ocean that were thought to be so abundant that it was not worth restricting their use.[222] Data are similarly plenteous.

Yet the overabundance of data is actually part of the problem. Making a commons of pooled personal data presents a modern update to the tragedy of the commons in which the harm is not depletion of the resource but rather the risk of multiplication of privacy and related harms from sharing, duplicating, and spreading the resource. Indeed, privacy is the preeminent justification today for limiting the sharing of big data. The risk of proliferating privacy harms looms as one of the most formidable objections against recognizing a right of public interest access to our pooled personal data.

The response to the objection cannot simply be to accept the status quo. Rather, we must determine some way to manage this contemporary configuration of the tragedy of the commons. This is fundamentally a question about the governance of common-pool resources.

In groundbreaking work that earned the Nobel Memorial Prize in Economic Sciences, Elinor Ostrom delineated eight important principles for governing common-pool resources.[223] For the electronic data context, the list can be winnowed to five informative insights. First, a common-pool resource is not necessarily open to all, but who or what entities are qualified to access the resource must be clearly defined.[224] Second, entities with access and use rights must have recognized space to formulate rules of organization and monitoring without top-down exclusive governmental control.[225] While governmental rules can provide scaffolding, within that scaffolding there must be space and recognition for self-regulatory activities.[226] Third, there must be rapid access to low-cost, potentially informal conflict resolution mechanisms to address ambiguities and conflicts.[227] Fourth, there must be graduated sanctions to address violations of the operational rules.[228] Fifth, entities with access and use rights can supplement monitoring by formal legal government structures, which are more cumbersome and

---

222 *See* Rose, *supra* note 22, at 717–18 (discussing the "plenteous good" rationale).

223 ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION 90–102 (1990) [hereinafter OSTROM, GOVERNING THE COMMONS]; *see also, e.g.*, NAT'L RSCH. COUNCIL, THE DRAMA OF THE COMMONS 15 (2002).

224 OSTROM, GOVERNING THE COMMONS, *supra* note 223, at 90–92.

225 *Id.* at 93, 101.

226 *Id.* at 101.

227 *Id.* at 100.

228 *Id.* at 94–99.

potentially less nimble and savvy than participants in the shared resource.[229]

Ostrom's insights can inform how access and use rights for our pooled personal data should be governed. First, the right of the public to benefit from our common pool resource of data does not mean unfettered public access. Rather, access must be controlled through clear standards, definitions, and review processes for researchers with the training and capacity to utilize the data to contribute to public knowledge or produce other public benefits and provide data security. Second, the review and access procedures should incorporate rules of organization and monitoring that have arisen within the professional spheres of scientific research and business with a proven track record of successful regulation. Third, there must be provision for low-cost, potentially informal mechanisms for addressing appeals of denials of access and to clarify the rules and standards surrounding data access. Fourth, safe harbors for providing public interest access does not mean no sanctions at all. Rather, sanctions should be graduated for both researchers and businesses to give incentives for observing the rules without altogether chilling effective data sharing. Fifth, in the place of cumbersome governmental structures, review boards drawing expertise from industry and nonprofit organizations can provide mutual monitoring and conflict resolution.

B. Controlled-Access Strategies for Privacy Protection

Fortunately, time-tested models for access to sensitive data for research and harm prevention already exist and incorporate many of the above principles. These models follow what I term a controlled-access approach.[230] Controlled access means sharing data with professionals who are trained and bounded by professional ethics and institutional review boards to safeguard human subjects against privacy and other harms.[231] The controlled-access approach draws from the

---

[229] *Id.* at 94.

[230] FAN, CAMERA POWER, *supra* note 3, at 195–204 (discussing controlled access for police-worn body camera data to protect against privacy harms while facilitating violence prevention and civil rights protection research); Mary D. Fan, *Private Data, Public Safety: A Bounded Access Model of Disclosure*, 94 N.C. L. REV. 161, 198–203 (2015) (arguing for bounded access to privately held data of public concern and the merits of this model over mandated disclosure regimes).

[231] *See, e.g.*, COMM. ON STRATEGIES FOR RESPONSIBLE SHARING OF CLINICAL TRIAL DATA, BD. ON HEALTH SCIS. POL'Y, INST. OF MED. OF THE NAT'L ACADS., SHARING CLINICAL TRIAL DATA: MAXIMIZING BENEFITS, MINIMIZING RISK 139–58 (2015) (defining "controlled access" as "any arrangement whereby data sharers place certain restrictions on access to or conditions of use of data" and giving examples of possible conditions); John A. Robertson, *The Law of Institutional Review Boards*, 26 UCLA L. REV. 484, 485–94 (1979)

time-tested procedures and safeguards that have advanced knowledge, science, medicine, and population health.[232]

1. Safeguarding Our Most Sensitive Data

Medical and epidemiological research draw on large volumes of highly private and protected health data to protect the public and detect diseases and risk factors.[233] Pooling and sharing data for disease surveillance is a longstanding practice dating back to the nineteenth century.[234] Without the ability to analyze sensitive health data, trying to detect threats to population health would be laboring "in the darkness of ignorance."[235]

The data that health scientists analyze are some of the most highly protected and sensitive categories of information recognized and regulated today, subject to the stringent standards of the Health Insurance Portability and Accountability Act of 1996 (HIPAA).[236] HIPAA provides strong privacy protections, reflecting the cultural view that health information is deeply private and disclosure potentially extremely harmful.[237] Yet HIPAA also permits disclosure of data for research and public health purposes—even extremely sensitive health data with identifiers.[238]

The model for access to the valuable sensitive information is shaped by a history of cautionary horror tales that have led to the

(discussing the rise of institutional review board requirements for researchers and their institutions).

[232] *See* Matthew R. Sydes, Anthony L. Johnson, Sarah K. Meredith, Mary Rauchenberger, Annabelle South & Mahesh K.B. Parmar, *Sharing Data from Clinical Trials: The Rationale for a Controlled Access Approach*, TRIALS, Mar. 23, 2015, at 1, 5.

[233] Lawrence O. Gostin, Scott Burriss & Zita Lazzarini, *The Law and the Public's Health: A Study of Infectious Disease Law in the United States*, 99 COLUM. L. REV. 59, 82 (1999) (explaining that researchers and policymakers can collect vital statistics—including information about births and deaths, the ethnic and racial makeup of communities, and risk factors for ill health—after state authorization to assess health needs or risks in a population which can, if timely published, facilitate legislative action); Scott F. Wetterhall & Eric K. Noji, *Surveillance and Epidemiology* ("Public health surveillance is the cornerstone of epidemiology."), *in* THE PUBLIC HEALTH CONSEQUENCES OF DISASTERS 37 (Eric K. Noji ed., 1997).

[234] For histories, see, for example, ARCHON FUNG, MARY GRAHAM & DAVID WEIL, FULL DISCLOSURE: THE POLITICS, PERILS AND PROMISE OF TARGETED TRANSPARENCY 142, 183–215 (2007); Denise Koo & Scott F. Wetterhall, *History and Current Status of the National Notifiable Diseases Surveillance System*, 2 J. PUB. HEALTH MGMT. & PRAC. 4, 4–8 (1996).

[235] *See* John W. Trask, *Public Health Administration: Its Dependence Upon Reports of Cases of Sickness*, 28 PUB. HEALTH REP. 1, 2 (1913).

[236] Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified in scattered sections of 26 U.S.C., 28 U.S.C., and 42 U.S.C.).

[237] Roberta B. Ness, *Influence of the HIPAA Privacy Rule on Health Research*, 298 J. AM. MED. ASS'N 2164, 2164–68 (2007).

[238] 45 C.F.R. §§ 164.502(a)(1), 164.512(i), 164.514(e) (2020).

development of internal professional ethical standards and external laws for the use of data from human subjects.[239] After World War II, among the atrocities adjudicated at Nuremberg were brutal medical experiments by Nazi doctors on prisoners.[240] In the scientific realm, the drive to never again sink to such depravity resulted in the landmark Nuremberg Code, hailed as "the most important document in the history of the ethics of medical research."[241] Key requirements in the Code include voluntary informed consent, a beneficial-good requirement such that the "degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment," and the right to withdraw from experiments without repercussions.[242]

The laws and customs of research using sensitive human subject data underwent further development in the United States, where several infamous cases led to the National Research Act of 1974.[243] The federal law created the institutional review board (IRB) system to regulate research with human subjects.[244] The development of protections was spurred by controversial cases of research on people without their knowledge or informed consent, such as the 1962 study of thalidomide as a treatment for pregnancy symptoms without informing the subjects they were taking an experimental drug—which caused birth defects—and the infamous 1932–1972 Tuskegee syphilis study, which left syphilis untreated in Black male subjects even after penicillin became the standard of care.[245] The debacles decried in a 1966 *New England Journal of Medicine* article[246] spurred Congress to act and also resulted in the influential 1978 Belmont Report, drafted by an expert commission charged by Congress.[247]

Evaluating whether a proposal produces sufficient benefits to outweigh potential harms also has a long tradition in the health sci-

---

239 *See* THE BELMONT REPORT, *supra* note 132.

240 *See, e.g.*, Jay Katz, *The Nuremberg Code and the Nuremberg Trial: A Reappraisal*, 276 J. AM. MED. ASS'N 1662 (1996); Michael A. Grodin, *Legacies of Nuremberg: Medical Ethics and Human Rights*, 276 J. AM. MED. ASS'N 1682 (1996).

241 Evelyne Shuster, *Fifty Years Later: The Significance of the Nuremberg Code*, 337 NEW ENG. J. MED. 1436, 1436 (1997).

242 2 TRIALS OF WAR CRIMINALS BEFORE THE NUREMBERG MILITARY TRIBUNALS: "THE MEDICAL CASE," "THE MILCH CASE" 181–82 (1949), https://www.loc.gov/rr/frd/Military_Law/pdf/NT_war-criminals_Vol-II.pdf [hereinafter *Permissible Medical Experiments*].

243 National Research Act, Pub. L. No. 93-348, 88 Stat. 342 (1974).

244 For a summary, see ELIZABETH A. BANKERT & ROBERT J. AMDUR, INSTITUTIONAL REVIEW BOARD: MEMBER HANDBOOK 7–16 (3d ed. 2011).

245 For a discussion, see Todd W. Rice, *The Historical, Ethical, and Legal Background of Human-Subjects Research*, 53 RESPIRATORY CARE 1325 (2008).

246 Henry Beecher, *Ethics and Clinical Research*, 274 NEW ENG. J. MED. 1354 (1966).

247 THE BELMONT REPORT, *supra* note 132.

ences. The 1947 Nuremberg Code mandated that the "risk to be taken should never exceed that determined by the humanitarian importance" of the reasons for the request.[248] The Belmont Report's influential conception of beneficence also entails maximizing the possible benefits while minimizing the possible harms.[249] Today an important criterion that IRBs apply is that the "[r]isks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result."[250]

As a result of law and internal professional ethical developments, scientists working with human subjects' data today regularly apply numerous safeguards to prevent harms to human subjects.[251] Federal law requires IRBs to ensure that the disclosure or use of human subjects' data "involves no more than a minimal risk to the privacy of individuals" and requires researchers to submit data protection plans with the following elements:

> (1) An adequate plan to protect the identifiers from improper use and disclosure;
> (2) An adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law; and
> (3) Adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research for which the use or disclosure of protected health information would be permitted.[252]

To receive IRB approval, researchers also must show that sensitive health data are necessary for the project, and that the project could not be practicably conducted without the information.[253] Sanctions for violations of human subjects' protections include a phalanx of professional, civil, and criminal penalties.[254] The right of controlled public interest access to our pooled data commons can interface with this well-developed system of safeguards. The procedures and protections already tried and proven in the health sciences show the feasibility of controlled access to our pool of personal data for public interest purposes while minimizing the risk of privacy and other harms.

---

248 *Permissible Medical Experiments*, *supra* note 242, at 182.
249 THE BELMONT REPORT, *supra* note 132, § B(2).
250 45 C.F.R. § 46.111(a)(2) (2020).
251 *E.g.*, *id.* §§ 46, 160, 162, 164; THE BELMONT REPORT, *supra* note 132.
252 45 C.F.R. § 164.512(i)(2)(ii)(A)(1)–(3).
253 *Id.* § 164.512(i)(2)(ii)(B)–(C).
254 *See, e.g.*, 42 U.S.C. § 1320d-5.

The appeal of a controlled-access model of data-sharing is evidenced by the European Commission's proposed Digital Service Act.[255] The proposal would require "very large online platforms," with an average monthly active user base of 45 million or more, to give data access to "vetted researchers" to investigate "systemic risks" posed by the operation and use of the platform's services.[256] To be "vetted," a researcher must "be affiliated with academic institutions, be independent from commercial interests, have proven records of expertise in the fields related to the risks investigated or related research methodologies, and shall commit and be in a capacity to preserve the specific data security and confidentiality requirements."[257] Limiting access to qualified researchers with the training and capacity to undertake research in the public interest and protect sensitive shared data strikes a balance that allows beneficial uses while reducing potential harms.

2. *Applying Controlled Access to Privately Held Consumer Big Data*

To effectively administer the proposed right of access to pooled consumer data, legislation can specify that a business must provide a mechanism for researchers to submit proposals and a review process that evaluates projects and teams to ensure (1) the project is likely to produce knowledge that benefits the public; (2) the proposed research team has the professional qualifications to carry out the project; and (3) the team has the appropriate safeguards and training to protect shared data and prevent privacy and related harms.

While framing language will vary in the process of negotiation and codification, the key aspects of the right of public interest access and use should include recognition that: (1) access to consumer data for nonprofit research in the public interest poses a different balance of benefits and risks to individuals and the public than commercial use of such data; (2) because valuable pools of personal data only exist because of the individual contributions of the public, the right of the public to benefit must be recognized; and (3) businesses must provide a mechanism for public interest access and use of pooled personal data that includes a review of proposals for merit and the qualifications of the research team to (i) conduct the proposed research, (ii) protect data that are shared, and (iii) adhere to prevailing ethical

255 *Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC*, COM (2020) 825 final, arts. 25(1), 26(1), 31 (Dec. 15, 2020).

256 *Id.*

257 *Id.* art. 31(4).

and institutional review standards for research with sensitive data involving human subjects.[258]

As for the composition of boards that get to decide whether a project sufficiently serves the public interest to justify access, current practices in the sciences also yield experience-tested approaches. The IRB or privacy board evaluates researcher qualifications and whether projects present the likelihood of yielding sufficient public benefits to justify the disclosure and use of highly private data.[259] As for the composition of these gatekeepers, federal law requires at least five members of sufficiently varying professional, demographic, cultural, and attitudinal backgrounds to have the competence to review the range of proposals received and promote confidence that the IRB understands, and is protecting, the welfare of human subjects in the community.[260]

Similar principles should apply to the composition of review boards for public interest access to privately held pooled personal data. The membership of the review boards for the sharing of private data must promote confidence that the board is protecting and promoting public wellbeing and has the professional competence to do so. The review boards may be situated externally, such as in a nonprofit educational institution, or sit in-house in a company. Regardless of where the structure sits, to ensure competence and promote confidence, the composition must include experts with the professional, nonpartisan qualifications to review the merits of research proposals and include a majority of independent members who are not mere corporate insiders. Corporate entities subject to such data-sharing requirements would be permitted—and indeed encouraged—to pool their resources where feasible, such as in supporting shared independent review boards with access to pooled data maintained with safeguards to protect commercially sensitive proprietary information from unintended uses. The pooling of review boards and data could reduce the burden on researchers from having to separately apply to numerous companies for similar types of consumer data.

Despite the reluctance to share data in the private sector, the growing recognition of the benefits of sharing information has led to

---

[258] The time-tested ethical standards and review mechanisms to ensure protection of human subjects, including prevention of privacy harms, are discussed, *supra*, in Section III.B.1.

[259] *See* 45 C.F.R. § 46.111(a)(2) (2020).

[260] 21 C.F.R. § 56.107(a) (2020) ("The IRB shall be sufficiently qualified through the experience and expertise of its members, and the diversity of the members, including consideration of race, gender, cultural backgrounds, and sensitivity to such issues as community attitudes, to promote respect for its advice and counsel in safeguarding . . . human subjects.").

the development of data-sharing platforms to facilitate access.[261] Such data-sharing platforms, coupled with a legitimate proposal review procedure, can reduce the inefficiencies of navigating patchworks of information held by different private entities.

Creating a right of public access does not need to vitiate the existing recognition of property protections in compiled data—and indeed can leverage and build upon that stable baseline. The successes of the open innovation movement show how providing property protections can also create a foundation for a thriving creative commons.[262] Software copyright holders launched a flourishing creative commons by granting access to their intellectual property via General Public Licenses that required users also to commit to keeping innovations resulting from access to the intellectual property via the licenses free and open.[263] The result is "a virtuous cycle" in which improvements are also given back to the commons.[264]

Today a sophisticated Creative Commons for copyrighted works has emerged in which authors of works can choose a menu of rights they wish to reserve and terms on which they will share their intellectual property.[265] Patent holders have joined the open innovation party, using patent pledges to the public to spur the spread of technology platforms and standards such as Wi-Fi wireless network protocols or the 4G LTE wireless communication standard, advancing shared interests and collective goals.[266] The sharing of patented material typically proceeds under licenses such as the FRAND commitment, an acronym referring to terms that are "fair, reasonable and non-discriminatory."[267]

The implementation of a right to access also can build from the current baseline of property protections for data compiled by companies. A right of access does not mean a free-for-all grab by all comers. Businesses have the incentives and expertise to evaluate the necessary data protection safeguards that researchers must have in place to prevent privacy and other harms. Access can proceed on a license model

---

[261] *See* Heiko Richter & Peter R. Slowinski, *The Data Sharing Economy: On the Emergence of Information Sharing Intermediaries*, 50 INT'L REV. INTELL. PROP. & COMPETITION L. 4, 9–10 (2019).

[262] *See, e.g.*, JAMES BOYLE, THE PUBLIC DOMAIN 166–84 (2008) (offering an overview of the creation of the creative commons by creators of free and open software).

[263] *Id.* at 167–68.

[264] *Id.* at 176.

[265] *Id.* at 181.

[266] *See* Jorge L. Contreras, *Patent Pledges*, 47 ARIZ. ST. L.J. 543, 549 (2015).

[267] *See, e.g.*, Mark A. Lemley, *Intellectual Property Rights and Standard-Setting Organizations*, 90 CALIF. L. REV. 1889, 1906 (2002) (reporting on the prevalence of FRAND terms).

with requirements for data protection, knowledge sharing, and a demonstrated likelihood of benefitting the public.

C. Safe Harbors and Regulatory Sandboxes for Public Interest Sharing

Currently there are considerable risks and scant incentives for companies to share data with external researchers for public interest purposes.[268] Consider again the difficult negotiations for the release of data on web pages shared by Facebook users.[269] In 2019, prominent scholars and pioneers in industry-academic partnerships, Gary King and Nathaniel Persily, announced an innovative partnership model with Facebook for the release of what would be one of the largest social science datasets ever available.[270] They thought it would take about two months of work until the release of the data—instead, it took twenty months because of difficult negotiations over how to release data useable for research while satisfying privacy laws.[271]

One of the major challenges that shadowed and constricted the endeavor was the entry into effect of the GDPR—and the severe sanctions it imposes, as discussed in Section I.A.[272] Reluctance to share was intensified by the $5 billion fine and extensive Federal Trade Commission oversight that Facebook incurred after the Cambridge Analytica data breach scandal, in which a political consulting firm obtained the data of tens of millions of Facebook users in an effort to learn how to manipulate potential voters.[273] The data Facebook ultimately released had details obscured using differential privacy techniques that perturb data through censoring and noise, making the studies of individuals and small groups infeasible.[274]

---

268 *See* HARRIS & SHARMA, *supra* note 189, at 16 ("There is currently no high-profile governmental or private sector mechanism that provides a social incentive for companies to share data in support of academic projects that benefit society, but may not be closely aligned with a company's research interests or mission.").

269 *See* King & Persily, *Facebook URLs Dataset*, *supra* note 9 (discussing challenges).

270 *See* King & Persily, *A New Model*, *supra* note 191, at 705–06.

271 King & Persily, *Facebook URLs Dataset*, *supra* note 9.

272 *See supra* Section I.A; *see also* King & Persily, *Facebook URLs Dataset*, *supra* note 9 ("The greatest barrier we have faced concerned Facebook's interpretation of the relevant privacy restrictions contained in the General Data Protection Regulation (GDPR) from the European Union and the consent decree they operate under with the Federal Trade Commission.").

273 Cecilia Kang, *$5 Billion Fine for Facebook on User Data*, N.Y. TIMES, July 13, 2019, at A1; *see also* King & Persily, *Facebook URLs Dataset*, *supra* note 9 ("However, we are not the ones who have had to pay a five billion dollar fine in the wake of the Cambridge Analytica scandal, and we would not be on the hook if our legal interpretation did not win the day in court or with regulators.").

274 *See* King & Persily, *Facebook URLs Dataset*, *supra* note 9 ("The privacy protective procedures instituted mean that researchers will not be able to learn about any individual

The difficult negotiations—shadowed by privacy laws—led King, Persily, and Harvard's Social Science One, which is dedicated to building industry-academic research partnerships for science, to call for "safe harbors specifically for research on social media data."[275] Going further, these pioneers of public-private sector data-sharing also dreamed of a "mandate that these companies share privacy-protected data with independent academics under a broad regulatory regime aimed at transparency."[276] This Section presents two proposals to create the incentives and fertile environment for sharing privately controlled consumer data: statutory safe harbors for public interest sharing and regulatory sandboxes.

1. Statutory Safe Harbors

Statutory safe harbors from sanctions arising from data sharing for public interest purposes can help redress this imbalance in risks and incentives. The law frequently uses safe harbors from liability to incentivize desired behavior or foster the development of new innovations.[277] For example, antitrust law enforcement agencies have created

---

or their actions, and small groups will also be obscured in the data which may make certain valid research questions impossible."); Liu, *supra* note 77, at 493–94 (explaining differential privacy techniques).

275 King & Persily, *Facebook URLs Dataset*, *supra* note 9.

276 *Id.*

277 *See, e.g.*, 35 U.S.C. § 271(e)(1) (providing a safe harbor from infringement liability for making, using, importing, selling, or offering to sell "a patented invention . . . which is primarily manufactured using recombinant DNA, recombinant RNA, hybridoma technology, or other processes involving site specific genetic manipulation techniques"); Embrex, Inc. v. Serv. Eng'g Corp., 216 F.3d 1343, 1349 (Fed. Cir. 2000) (recognizing a limited "experimental use" exception to patent infringement); *see also infra* notes 278–82 (discussing informational safe harbors in antitrust law and the Digital Millennium Copyright Act's safe harbor provisions); Bidwell v. Univ. Med. Ctr., 685 F.3d 613, 615 (6th Cir. 2012) (noting new Department of Labor regulations that created safe harbors from fiduciary liability for pension plan administrators who invested in certain types of riskier short-term investments with a potential for higher yields to create incentives for such investments); ALAN R. BROMBERG, LEWIS D. LOWENFELS & MICHAEL J. SULLIVAN, BROMBERG & LOWENFELS ON SECURITIES FRAUD § 5:276 (2d ed. 2021) (noting "the 1995 enactment of the Private Securities Litigation Reform Act ('PSLRA'), which contains a safe harbor from liability for many types of forward-looking statements if they are accompanied by 'meaningful cautionary statements,' has lent a new incentive to risk disclosure under the federal securities laws" (internal citations omitted)); Judson D. Stelter, Note, *The IRS' Classification Settlement Program: Is It an Adequate Tool to Relieve Taxpayer Burden for Small Businesses That Have Misclassified Workers as Independent Contractors?*, 56 CLEV. ST. L. REV. 451, 461 (2008) (noting the safe harbor in the tax code "for employers who have misclassified employees as independent contractors . . . so long as the employer meets three requirements: reporting consistency, substantive consistency, and a reasonable basis for the classification"). *See generally* Robert Gatter, *Walking the Talk of Trust in Human Subjects Research: The Challenge of Regulating Financial Conflicts of Interest*, 52 EMORY L.J. 327, 397 (2003) (discussing the advantages of creating safe harbors instead of a list of prohibitions to incentivize desired behaviors).

"safety zones" for the sharing of information between business competitors that will not incur liability to encourage certain kinds of information exchanges, for example to coordinate cybersecurity.[278] Another antitrust safety zone for competitor communication of information is recognized to facilitate collaborative research and development efforts under certain kinds of innovation market conditions.[279]

Famously—or infamously to scholars who think the training wheels should come off—statutory safe harbors also fostered the growth of the Internet through protections for Internet Service Providers (ISPs).[280] The Digital Millennium Copyright Act (DMCA), passed in 1998, immunized ISPs from monetary damages for providing several key services, including: (1) Internet access, (2) temporary storage or caching of data, (3) passively storing or hosting user materials, and (4) giving users location tools, such as linking to content on various websites.[281] Public or nonprofit institutions of higher education that act as ISPs also received a safe harbor from liability for copyright-infringing acts by faculty members and graduate students.[282] Whether and when safe harbors like these outgrow their usefulness and should sunset is a different debate. The point is that safe harbors can be useful in encouraging socially beneficial innovation and behaviors.

Selling data to third parties for commercial exploitation is fundamentally different than sharing data for public interest purposes to researchers subject to the external review and ethical safeguards used

---

[278] Michael Bloom, *Information Exchange: Be Reasonable*, FED. TRADE COMM'N (Dec. 11, 2014), https://www.ftc.gov/news-events/blogs/competition-matters/2014/12/information-exchange-be-reasonable; FED. TRADE COMM'N & U.S. DEP'T OF JUSTICE, ANTITRUST POLICY STATEMENT ON SHARING OF CYBERSECURITY INFORMATION 3–9 (2014), https://www.ftc.gov/public-statements/2014/04/department-justice-federal-trade-commission-antitrust-policy-statement.

[279] FED. TRADE COMM'N & U.S. DEP'T OF JUSTICE, ANTITRUST GUIDELINES FOR COLLABORATIONS AMONG COMPETITORS 26–27 (2000), https://www.ftc.gov/sites/default/files/documents/public_events/joint-venture-hearings-antitrust-guidelines-collaboration-among-competitors/ftcdojguidelines-2.pdf (creating a research and development communication safe harbor "in an innovation market where three or more independently controlled research efforts in addition to those of the collaboration possess the required specialized assets or characteristics and the incentive to engage in R&D that is a close substitute for the R&D activity of the collaboration").

[280] For a discussion, see, for example, Nicholas W. Bramble, *Safe Harbors and the National Information Infrastructure*, 64 HASTINGS L.J. 325, 332–43, 350–63 (2013) (explaining safe harbors and the novel ways in which they promote the growth of speech infrastructure); Edward Lee, *Decoding the DMCA Safe Harbors*, 32 COLUM. J.L. & ARTS 233, 235–38 (2009) (explaining how safe harbors interact with vicarious liability); Mark A. Lemley, *Rationalizing Internet Safe Harbors*, 6 J. TELECOMM. & HIGH TECH. L. 101, 104–05 (2007) (discussing safe harbors pertaining to internet intermediaries).

[281] Digital Millennium Copyright Act, 17 U.S.C. § 512(a)–(e).

[282] *Id.* § 512(e).

in scientific research. The rules and risks of liability for philanthropic data sharing should also be different to incentivize socially desirable conduct. While the details and statutory language will vary with political negotiations and compromises to secure passage, the key elements of a safe harbor provision should: (1) distinguish between commercial use of pooled personal data and sharing with external researchers from the nonprofit sector for public interest purposes, and (2) curtail and limit the damages and fines that are otherwise applicable for infringements, such as data breaches, that arise from transfers of data to researchers for public interest purposes so long as there was a good-faith, reasonable review of the researchers' qualifications and capabilities to provide data security.

As discussed in Section III.B.1, researchers have extensive experience and review processes for protecting and using some of the most sensitive data available—the health information of human subjects.[283] The risk of sanctions from one-size-fits-all liability rules should not chill data sharing, or force release of data so censored and perturbed that much important utility is lost.[284]

2. *Regulatory Sandboxes*

Another approach would draw on a more recent policy development to encourage innovation by private firms—regulatory sandboxes.[285] According to the first governmental authority to deploy the concept, the UK's Financial Conduct Authority, the fundamental idea behind a regulatory sandbox is to create a safe space for innovation and experimentation by relaxing regulations and punishments when businesses or organizations test new approaches, products, or technologies.[286] Entities applying to regulators for regulatory sandbox treatment have to ensure they apply safeguards to protect consumers or other protected persons or entities from the adverse effects of the experimentation.[287]

---

[283] *See supra* Section III.B.1.

[284] *See supra* note 274 and accompanying text.

[285] *See, e.g.*, Ivo Jenik & Kate Lauer, Regulatory Sandboxes and Financial Inclusion 1 (Oct. 2017) (working paper), https://www.cgap.org/sites/default/files/researches/documents/Working-Paper-Regulatory-Sandboxes-Oct-2017.pdf ("A regulatory sandbox is a framework set up by a financial sector regulator to allow small scale, live testing of innovations by private firms in a controlled environment (operating under a special exemption, allowance, or other limited, time-bound exception) under the regulator's supervision.").

[286] Aaron Martin & Giulia Balestra, *Using Regulatory Sandboxes to Support Responsible Innovation in the Humanitarian Sector*, 10 GLOB. POL'Y 733, 733–34 (2019).

[287] *See* Dirk A. Zetzsche, Ross P. Buckley, Janos N. Barberis & Douglas W. Arner, *Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation*, 23 FORDHAM J. CORP. & FIN. L. 31, 64 (2017) (describing the basics of setting up a regulatory sandbox).

Thus far, regulatory sandboxes are most known in the United States as a strategy by financial industry regulators to encourage innovation in financial technology.[288] The U.S. Consumer Financial Protection Bureau (CFPB) was the first domestic regulatory authority to formally announce plans to deploy a regulatory sandbox for companies to test new strategies for consumer disclosures.[289] Other U.S. regulatory agencies, such as the Securities and Exchange Commission (SEC) and the Federal Deposit Insurance Corporation (FDIC), have announced plans to join a cross-border regulatory sandbox effort.[290]

Globally, regulatory sandboxes are a growing phenomenon extending to sectors beyond the financial industry.[291] Recently, public-sector scholars and officials have proposed appropriating the private-sector use of regulatory sandboxes typically used to foster business innovation to serve humanitarian goals.[292] For example, regulatory sandboxes can help nongovernmental organizations and international organizations address regulatory challenges and uncertainties that inhibit partnerships to share sensitive data to respond to refugee crises in a timely, effective, and coordinated manner.[293] Regulatory sandboxes may relieve organizations of certain privacy restrictions for limited approved purposes where the benefits outweigh the risks of regulatory relaxation.[294]

The concept of a regulatory sandbox also can be used to encourage businesses to share consumer big data for limited public benefit purposes, with controlled-access protections in place to reduce the risks of harm. In this regulatory sandbox constructed for the purpose of facilitating research to benefit the public, companies that share consumer big data with such protections in place would not be subject to the onerous privacy regulations and potentially severe pun-

---

288 *See* Hilary J. Allen, *Regulatory Sandboxes*, 87 GEO. WASH. L. REV. 579 (2019).

289 *CFPB Office of Innovation Proposes "Disclosure Sandbox" for Companies to Test New Ways to Inform Consumers*, CONSUMER FIN. PROT. BUREAU (Sept. 13, 2018), https://www.consumerfinance.gov/about-us/blog/cfpb-office-innovation-proposes-disclosure-sandbox-companies-test-new-ways-inform-consumers.

290 Press Release, U.S. Sec. & Exch. Comm'n, U.S. Financial Regulatory Agencies Join the Global Financial Innovation Network (Oct. 24, 2019), https://www.sec.gov/news/press-release/2019-221.

291 *See, e.g.*, Martin & Balestra, *supra* note 286, at 735 (describing how regulatory sandboxes may be used by humanitarian actors).

292 *See* Jenik & Lauer, *supra* note 285, at 2; Martin & Balestra, *supra* note 286, at 733–34.

293 Martin & Balestra, *supra* note 286, at 733–34.

294 *See, e.g.*, NSW GOV'T (AUST.), BRINGING BIG IDEAS TO LIFE: NSW INNOVATION STRATEGY 7 (2015), https://www.acs.org.au/content/dam/acs/acs-documents/NSW_Government_Innovation_Strategy_Document.pdf (relaxing data privacy regulations and related potential barriers under certain circumstances to foster innovation).

ishments of regimes such as the GDPR.[295] This would help alleviate the pressure on companies to release data that are so masked and distorted that their research value is severely reduced.[296]

Sharing that fosters knowledge creation requires useable data. For example, free and open-source software made available under the General Public License granted public access "to the human-readable 'source code' rather than just the inscrutable 'machine code,'" to permit innovators to "understand, tinker, and modify."[297] With safe harbors or regulatory sandboxes in place, the unfortunate incentives for companies to withhold or obscure data from researchers are corrected.

Sharing of consumer data for public benefit purposes is generally nonrivalrous with use by private companies for profitmaking—for example, the information remains lucrative for customization of advertising even if it is shared with safeguards to researchers to prevent substance abuse, suicides, or accidental overdoses.[298] Indeed, it is in the interest of companies, as well as the public, to show potentially beneficial uses of the deluge of data, particularly in times of growing consternation and hostility toward the aggregation of information about us.[299] This is an opportune time to recognize the right of the public to benefit from our valuable pooled consumer big data. Protections that draw on controlled-access models and incentives to foster innovation can create the conditions to better distribute the benefits of our pooled personal data while protecting against privacy harms.

---

[295] For a discussion of the penalties and perverse incentives they create, see *supra* Section I.A and text accompanying notes 272–76.

[296] *See supra* Section I.A and notes 271–75 and accompanying text.

[297] BOYLE, *supra* note 262, at 167.

[298] *See, e.g.*, Marie C. Baca, *What You Do on the Internet Is Worth a Lot. Exactly How Much, Nobody Knows*, WASH. POST (Oct. 14, 2019, 7:00 AM), https://www.washingtonpost.com/technology/2019/10/14/what-you-do-internet-is-worth-lot-exactly-how-much-nobody-knows (discussing efforts to quantify the lucrative nature of consumer data to companies for uses such as targeted advertising and studying consumer behavior to better sell products); *Using Social Media to Better Understand, Prevent, and Treat Substance Abuse*, NAT'L INST. ON DRUG ABUSE (Oct. 16, 2014), https://archives.drugabuse.gov/news-events/news-releases/2014/10/using-social-media-to-better-understand-prevent-treat-substance-use ("Researchers can analyze social media interactions to gain insights into patterns of use, risk factors, and behaviors associated with substance use.").

[299] *See, e.g.*, Morey et al., *supra* note 37 ("If companies understand how much data is worth to consumers, they can offer commensurate value in return for it. Making the exchange transparent will be increasingly important in building trust.").

Conclusion

There is growing recognition that big data is a natural resource for which wise stewardship is required.[300] Part of the wise stewardship is ensuring that the public also benefits from pooled personal data currently largely held by businesses for commercial gain. Distributing the benefits of big data derived from our personal information and internet trails to the public is in the interest of businesses as well as the public, as it would address widespread concern that while people bear the harms of consumer data collection and aggregation, they do not reap the benefits.[301]

The current slew of legislative proposals and enactments focused on individual privacy protections leaves unaddressed the larger issue of rights in our pooled personal data.[302] One's personal data is a drop in the ocean of the value and power of consumer big data. Granting greater rights of individual control does not answer the larger question of rights in the most valuable and powerful asset—our aggregated personal data.

This Article provides the foundation for recognizing the right of the public to benefit from pooled personal data currently controlled largely for private profit.[303] Opening the consumer big data commons for public benefit does not have to entail a tragedy of the commons in terms of privacy and related harms.[304] Insights from commons governance, open innovation, regulatory advances, and well-established procedures in human subjects protection can inform a model of controlled access and use of our pooled personal data to allow public participation in the benefits of the resource, while mitigating the risk of harm.[305]

---

300 *See, e.g.*, Antonio Neri, *We Should Treat Data as a Natural Resource. Here's Why*, World Econ. F. (Mar. 2, 2020), https://www.weforum.org/agenda/2020/03/we-should-treat-data-as-a-natural-resource-heres-why (advocating for treating data as a natural resource to be harnessed "to drive progress on some of the world's most intractable societal challenges"); Mike Smith, *Data Is the World's New Natural Resource*, IBM A/NZ Blog (Feb. 12, 2019), https://www.ibm.com/blogs/ibm-anz/data-is-the-worlds-new-natural-resource ("Data is the world's new natural resource, unleashed by the maturation of artificial intelligence, and holds the potential to generate economic wealth, health and social wellbeing. As with any natural resource, it must be looked after. Good data stewardship must be transparent and done for a purpose.").

301 *See supra* notes 138–40; *see also, e.g.*, Morey et al., *supra* note 37 (discussing the business interest in rebuilding consumer trust).

302 *See supra* Section I.A.

303 *See supra* Section II.B.

304 *See supra* Section III.A.

305 *See supra* Sections III.B–C.