ENGELBERG CENTER
on Innovation Law & Policy
NYU School of Law

APRIL 2021

Interoperability and Portability in the Wild

Lessons from the Data Sharing Practitioners Workshop

Gabriel Nicholas

Engelberg Center on Innovation Law & Policy

The Engelberg Center on Innovation Law & Policy at New York University School of Law provides a unique environment where scholars can examine the key drivers of innovation as well as the law and policy that best support innovation. By fostering interdisciplinary and collaborative research on innovation law and policy, the Engelberg Center attracts legal scholars and practitioners, technologists, economists, social scientists, physical scientists, historians, innovators, and industry experts who study the incentives that motivate innovators, how those incentives vary among creative endeavors, and the laws and policies that help or hinder them.

About the Author

Gabriel Nicholas is a Joint Research Fellow at the NYU School of Law Information Law Institute and the NYU Center for Cybersecurity, and a Fellow at the Engelberg Center on Innovation Law & Policy.

Acknowledgments

This workshop was held over two sessions on January 19 and 27, 2021. It was hosted by the Engelberg Center on Innovation Law & Policy at NYU Law. The workshop was organized and moderated by the author.

ACADEMIC REVIEW BOARD

Thank you to the Academic Review Board members who attended this workshop, gave feedback in its early stages of planning, and helped organize the event. The members of the board are as follows:

Avery Gardiner — General Counsel and Senior Fellow for Competition, Data, and Power at the Center for Democracy & Technology

Inge Graef — Associate Professor of Competition Law at Tilburg University

Nizan Packin — Associate Professor at the Law Department of the Zicklin School of Business at City University of New York Baruch College

Katherine Strandburg — Alfred B. Engelberg Professor of Law at New York University School of Law

Peter Swire — Elizabeth and Tommy Holder Chair and Professor of Law and Ethics at the Georgia Tech Scheller College of Business

Michael Weinberg — Executive Director of the Engelberg Center on Innovation Law & Policy at New York University School of Law

Thanks also to Ruby Mayer, Randy Milch, Kerry Sheehan, Kyle Wiens, and the Fellows at the Information Law Institute at NYU Law for their input and their help in organizing this workshop.

Contents

Executive Summary

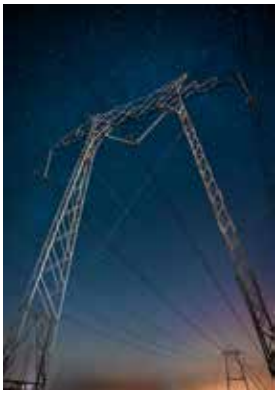
A new kind of tech company is on the rise. They are called *data users*, and they don't collect data themselves but rather allow consumers to bring their data over from elsewhere to realize its value in new ways. Data users offer novel benefits to consumers and improve competition and innovation in the tech sector at large. They often depend on data sharing regulation, particularly interoperability and portability, to make sure firms that control consumer data, called *data holders*, offer consumers the technical means to move their data to new services. Much ink has been spilled about how different sectoral regulators can balance the economic and consumer benefits of data sharing with the potential cybersecurity and privacy risks. However, these conversations tend to be dominated by the data holders, who have the resources to make themselves heard and the market incentives not to let consumers move their data. Data holders thus restrict data sharing, and data users, too siloed in their individual sectors to voice their shared concerns, are left to deal with the fallout.

The Engelberg Center on Innovation Law & Policy at NYU Law held the Data Sharing Practitioners Workshop to uncover the common difficulties data users have with data holders in how they allow consumers to share their data. The workshop brought together industry representatives from six sectors: healthcare, finance, energy, agriculture, automotive, and medical devices. Some of these industries have data sharing regulation already; others, particularly hardware-centric industries, have little to none. All face major barriers to portability and interoperability from data holders. This workshop summary paper is intended to give voice to the often-unheard but surprisingly consistent concerns and policy interests of data users. It is not meant to weigh these interests against those of data holders, nor to endorse any specific data sharing policy.

This workshop uncovered a few common strategies data holders employ to make it more difficult for consumers to bring their data to new services. Participants explained how data holders create poor user experiences for consumers sharing data with third parties and give preferential treatment to themselves and partners. When data users do receive data, it can be difficult for them to use because data holders often do not adopt common technical standards or document their proprietary standards. Together, these lead data users to find unsanctioned ways for consumers to bring their data over, such as screenscraping and reverse engineering, which create new privacy and security vulnerabilities for consumers and new opportunities for data holders to discriminate access.

Workshop participants also noted some common policy misconceptions that have let poor data sharing practices to proliferate. They questioned the common wisdom that competition between data holders naturally leads to improved data sharing practices. They also generally believed that debates over data ownership distract from more important questions about who has access to data and how. Finally, participants criticized policymakers for siloing conversations about data sharing from conversations about privacy and security. Workshop participants expressed interest in a few alternative approaches to improving data sharing policy: empowering uncaptured standards bodies, permitting alternative data access methods, and improving the consumer experience of sharing data.





I. Introduction

As the American economy matures into a data economy, firms are finding new ways to bring consumers value from their data. The companies discovering the most innovative uses for data, though, are not always the ones collecting it. Today, there is a burgeoning ecosystem of companies that let users bring their data in from elsewhere to realize its value in new ways. These companies use data from banks to build budgeting tools,[1] combat exorbitant fees,[2] and bring credit access to the unbanked.[3] They use data from utilities to help consumers use less energy when the grid is under duress[4] and offset their electricity usage with clean energy.[5] And they allow consumers to get more use out of their smart devices, whether by helping cars drive themselves[6] or helping patients gain health insights from their medical devices.[7] Many of these companies provide hope for competition and innovation in otherwise concentrated and entrenched industries.

Technically speaking, firms that collect data about consumers and hold onto data they provide are gatekeepers — they determine whether those consumers can bring their data elsewhere, where they can bring it, and how. Data holders tend to be incentivized by the market to restrict the data they allow consumers to transfer. Therefore, data users often depend on data sharing regulation, particularly portability and interoperability rules, to make sure consumers have the capacity to bring data over to their services. However, data sharing requirements can raise privacy concerns about how third parties might use the data they receive and cybersecurity threats about how available data pathways could be leveraged for unauthorized access. When data holders let consumers move their data, they risk being held legally and politically liable for the ways those who receive the data misuse it.

1 *E.g.* Mint, Yodlee, Quicken.
2 *E.g.* Cushion.ai.
3 *See* Ariadne Plaitakis & Stefan Staschen, *Open Banking: How To Design For Financial Inclusion*, Consultative Group to Assist the Poor (Oct. 2020), https://www.cgap.org/research/publication/open-banking-how-design-financial-inclusion
4 *E.g.* OhmConnect.
5 *E.g.* Arcadia.
6 *E.g.* Comma.ai.
7 *E.g.* Nightscout, OpenAPS.

In regulatory conversations about how to strike a balance between the costs and benefits of data sharing, policymakers have tended to give more weight to the concerns of data holders than those of data users. In part, this is because data holders often have more market power and political weight to push their agendas. However, it is also because data users are siloed within their own industries. While data holders have the rhetorically powerful North Stars of cybersecurity, privacy, and liability, data users do not have a shared language to raise concerns about the competitive and innovative harms poor data sharing practices can have on consumers and the economy at large.

The Engelberg Center on Innovation Law & Policy at NYU Law held the Data Sharing Practitioners Workshop to give voice to the shared challenges faced by data users across different sectors. Over two days, it brought together more than a dozen industry practitioners from six sectors to find out what barriers they encounter to consumer-permissioned data access and how they would like policymakers to help alleviate those barriers. Workshop participants had backgrounds in healthcare, finance, energy, agriculture, automotive, and medical devices. Despite coming from a wide range of regulatory and market environments, participants found they faced many of the same problems.

This paper organizes and summarizes the comments made by workshop participants. The goal here is not to offer policy prescriptions about how data sharing regulation should be designed, nor to balance the views of data users against those of data holders. Rather, it is to surface the concerns and policy interests of data users, a group that is traditionally disunited and underrepresented in policy conversations.

(A sidenote: The workshop did not include any firms that buy user data, such as ad firms, or otherwise use data that consumers do not knowingly transfer themselves. It also did not include representatives from industries like social media, where consumers derive most of the benefit from network effects, not value created by the data user.)

Part II of this paper defines some basic terms, describes the workshop, and gives high-level backgrounds on the market and regulatory dynamics for data sharing in each of the six sectors. Part III highlights common challenges participants face in their dealings with data holders. Part IV discusses shared concerns participants had about the effectiveness of different approaches to data sharing policy. Part V concludes and gestures to new approaches to regulation participants were interested in exploring.

II. Background

Definitions

Any industry that uses networked communication has some concept of sharing data; unfortunately, each has its own vocabulary to describe it. Despite valiant efforts, scholars and policymakers have failed to converge on common definitions for even the most basic terms, such as "portability" and "interoperability." Summarizing the results of a multidisciplinary, multi-sectoral workshop magnifies the definitional problem — a tractor engineer and a microfinancial data scientist mean very different things when they talk about "open access to data," but a shared vocabulary is necessary to compare their problems. This paper thus uses its own fairly general, technologically agnostic terminology as detailed below:

A **data holder** is an entity that controls access to a system that has data provided by or about a consumer. Different industries call this the provider, transferring entity, or in the case of hardware, original equipment manufacturer (OEM).

A **data user** is an entity that uses or is interested in using a consumer's existing data, with their permission, as part of a new product or service. Different industries call this the requester, receiving entity, or in the case of hardware, aftermarket vendor.

Data sharing is the process whereby a data holder allows a consumer to give a third party access to their data. This definition includes methods such as one-off export portability and interoperability. Data sharing can occur either through a system explicitly designed to facilitate the transfer of data, such as an application programming interface (API), or through means unsanctioned by the data holder, such as screenscraping. It is important to note that although it is possible to access a consumer's data without their explicit permission (e.g., through data brokers), such methods are excluded from this definition of "data sharing."[8]

8 Legal scholars often criticize the notice and consent paradigm for putting too much of a congnitive burden on consumers. These concerns are important for policymakers to consider but are also outside the scope of this paper. *See* Joel R. Reidenberg et al., *Privacy Harms and the Effectiveness of the Notice and Choice Framework*, 11 I/S: J.L. & POL'Y. FOR INFO. SOC'Y. 485 (2015); Jennifer M. Urban & Chris Jay Hoofnagle, *The Privacy Pragmatic as Privacy Vulnerable*, Berkeley Public Law Research Paper No. 2514381 (2014), http://ssrn.com/abstract=2514381

Workshop Description

The Data Sharing Practitioners Workshop was hosted by the Engelberg Center on Innovation Law & Policy at NYU Law on January 19 and 27, 2021. It brought industry experts who use shared data to build new products together with legal scholars and policy experts interested in how data sharing can more broadly improve competition and innovation. The workshop was divided into two sessions. The first focused on Internet of Things (IoT) and included data users from the agriculture, automotive, and medical device sectors. The second focused on non-IoT industries that already have some sort of data sharing regulation and included data users from the healthcare, finance, and energy sectors. The workshops were invitation-only, held over Zoom, and run according to modified Chatham House Rules style — that is, participants agreed to be identified by industry for this paper but not by name or affiliation. Researchers attended both workshops and six researchers were part of an academic board of advisors (see front matter for details). The IoT and non-IoT workshops had twelve and eleven attendees respectively.

Industry Backgrounds

Before delving into the workshop findings, it is important to give background context on the unique data sharing environments of each participant. Data sharing in any one of the six sectors represented is a vast field of study unto itself. This background section only scratches the surface of each and attempts to give just enough information on the market and regulatory conditions of each sector for readers to understand the rest of the paper. Each summary will touch on what data is relevant, who holds it, who is interested in using it, how the flow of data is regulated, and what potential consumer harms the flow of data raises. Background summaries will focus on the issues most relevant to workshop participants.

HEALTHCARE

Due to decades of regulatory carrots and sticks, most data collected by health providers on patients is digitized and stored online in what are called *electronic health records* (EHRs). Individual EHRs make it easier for healthcare providers to share and analyze data. En masse, EHRs allow for data analysis that can be used for everything from tracking pandemics to targeting pharmaceutical advertisements. The Health Insurance Portability and Accountability Act (HIPAA) gives patients a "right of access" to their data and allows them to send it where they please,[9] but at a technical level, EHR vendors determine which data patients can send and to whom.[10]

9 45 C.F.R. § 164.524

10 *See, generally* S. Trent Rosenbloom et al., *Updating HIPAA for the Electronic Medical Record Era*, 26 J. Amer. Med. Inform. Assoc., 1115-1119 (2019), https://academic.oup.com/jamia/article/26/10/1115/5544256

EHR vendors sometimes withhold data from consumers and data users. Patient data protected under HIPAA is subject to stringent privacy and security requirements, and EHR vendors express interest in protecting patients by making sure they do not bring data to non-HIPAA-compliant entities.[11] Data sharing also creates opportunities for health data aggregation, which can be used to re-identify de-identified data and potentially discriminate access.[12]

Data users claim that these privacy and security issues are not always raised in good faith, and recently, as part of the 21st Century Cures Act, regulators started to require EHR vendors to make more information interoperable and easily available.[13] The rule prevents certain methods of information blocking, requires EHRs to adopt APIs, and makes EHRs follow guidance from specific standards bodies.

FINANCE

Financial institutions such as banks and credit card companies hold data on consumers' transactions (e.g., amount, date, description) and account information (e.g., account type, balance, routing number).[14] Data users who are interested in transaction data include money management tools and financial inclusion products that offer alternative forms of credit. Data users interested in account data include business payment solutions and tools to make it easier for consumers to navigate the financial system.

Unlike the EU, the UK, and many other countries, the US has no major federal laws requiring financial institutions to make their data available to third parties upon a consumer's request. The US has instead taken a market-based approach, allowing industry-led initiatives by groups such as The Clearing House and the Financial Data Exchange to help banks and fintech companies find middle ground on data sharing standards and best practices. Banks, however have been hesitant to adopt these measures and often severely limit the data available through permissioned data flows. As a result, data users usually access consumer data through screenscraping, whereby they enter a data holder's system through a consumer's account login credentials and programmatically take information from their webpage.[15]

Screenscraping can cause privacy and security vulnerabilities. Improperly stored user credentials can lead to worst-case-scenario breaches, and screenscraping makes it impossible to limit how much information a data user collects.[16] Unchecked financial data aggregation raises its own concerns. Data users like Plaid have become powerful industry players, and because they are largely unregulated in how they use data, open the door for potential consumer harm.[17]

---

11 *See* Lucia Savage, *To Combat 'Information Blocking,' Look To HIPAA*, Health Affairs (Aug. 24, 2017), https://www.healthaffairs.org/do/10.1377/hblog20170824.061636/full/

12 Khaled El Emam et al., *A Systematic Review of Re-Identification Attacks on Health Data*, 6(12) PLOS ONE (Apr. 2015), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0028071

13 21st Century Cures Act, Pub. L. No. 114-255, 130 Stat. 1033, 1033 (2016) (codified as amended at 42 U.S.C. § 201 (2016)).

14 For more information on these two categories of financial data, see UK Open Banking's definitions of "Account Servicing Payment Service Provider" (ASPSP) and "Payment Initiation Services Provider" (PISP). Open Banking, *The Open Banking Glossary*, https://www.openbanking.org.uk/about-us/glossary/ (accessed Mar. 13, 2021).

15 Nizan Packin, *Show Me the (Data About the) Money!*, Utah L. Rev. (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3620025

16 U.S. Department of the Treasury, *A Financial System That Creates Economic Opportunities: Nonbank Financials, Fintech, and Innovation* 34-35 (2018) https://home.treasury.gov/sites/default/files/2018-08/A-Financial-System-that-Creates-Economic-Opportunities---Nonbank-Financials-Fintech-and-Innovation.pdf

17 Packin, *supra* note 15.

ENERGY

Consumers have had access to data on their electricity usage ever since utilities began mailing bills to people's homes. Today, utilities collect most residential, commercial, and industrial energy usage data through smart meters.[18] Numerous third-party products use the data they collect to build everything from budgeting tools to demand response programs that help reduce energy demand during peak times to applications that aid the shift to renewable energy.[19] Utilities, however, whether owned by investors, a municipality, or cooperatively, have been slow to make it easy for consumers to share their data with third parties. As highly regulated, government-sanctioned monopolies, utilities tend not to build new infrastructure without being required to by law or guaranteed that they will recoup the costs.

Data sharing practices vary widely among utilities because each is subject to the rules of different state regulators, city councils, and utility boards. The closest thing America has to a unified energy data sharing system is Green Button, an energy data sharing standard created by the Department of Energy that has so far been adopted by five states.[20] However, this standard was designed in 2012 and data users argue that the data they thought would be useful then is different from the data that is actually useful now. What is more, utilities have generally not invested in making their Green Button data sharing systems easy to use or fix them quickly when they break. Some utilities also run their own smart meter app stores, but again, data users complain that these stores let utilities reserve certain data for themselves and preference their own apps.[21] Many data users therefore turn to screenscraping, raising similar issues to finance, though with somewhat less acute privacy concerns.

AGRICULTURE

Data collection has become a key part of the agriculture business. Large equipment manufacturers like John Deere have created lines of "smart tractors" and other sensor-laden equipment to collect metrics like soil quality, moisture, and seed placement to allow farmers to hyper-customize their growing strategies. This data-driven approach to farming is known as *precision agriculture*.[22]

The shift toward computerization and precision agriculture has allowed original equipment manufacturers to lock out aftermarket competitors, particularly tractor headers and tows. Manufacturers withhold information about their proprietary interfaces, forcing aftermarket companies to reverse engineer their systems. Manufacturers like John Deere justify their poor interoperability by raising privacy and cybersecurity concerns, particularly about geolocation data, despite the fact that many aftermarket manufacturers do not collect data with their products.

---

18 U.S. Energy Information Administration, *Frequently Asked Questions (FAQs) - How many smart meters are installed in the United States, and who has them?* https://www.eia.gov/tools/faqs/faq.php?id=108&t=3 (accessed Mar. 13, 2021).

19 *See* Mission:data, *Energy Data Portability* (Jan. 2019), https://static1.squarespace.com/static/52d5c817e4b062861277ea97/t/5c3a849b562fa75d70fd7953/1547338949271/Energy+Data+Portability.pdf

20 *Id.*

21 *See* Mission:data, *Digital Platform Regulation of Electric Utilities* (Jan. 2021), https://static1.squarespace.com/static/52d5c817e4b062861277ea97/t/5ff3b7a4dc8f8e0b711f94fc/1609807781400/Digital+Platform+Regulation.pdf

22 *See* Solon Barocas, Karen Levy, & Alexandra Mateescu, *Reap What You Sow? Automation, Information, and Economic Distribution on the Farm*, We Robot (University of Miami School of Law: 2019), https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/BarocasLevyMateescu_WeRobot.pdf

No US federal laws specifically govern the data practices of agriculture companies, but the American Farm Bureau Federation (FB), a major industry group, has attempted to help the industry regulate itself. Studies show that the majority of farmers are concerned about data access and ownership issues,[23] and the FB has tried to address this concern with its Agricultural Data Transparency Evaluator, a certification that includes portability, choice, and preventing anti-competitive practices as core tenets.[24] Farm advocacy groups, however have complained that this certification is a rubber stamp, pointing out that John Deere is certified despite having some of the worst data sharing practices in the industry.[25]

AUTOMOTIVE

Most every component of the modern car is in some way connected to a computer. Cars collect data on their internal parts and use data to manage every aspect of the car, from driving to navigation to the entertainment system.[26] Many third parties are interested in accessing vehicle data. Independent repair shops require access to in-car mechanical data to diagnose and fix problems. Other companies seek to use the data to build new products, such as car insurance[27] and self-driving tools.[28]

Car manufacturers make some data available through their onboard diagnostics systems, either through a port in the car or remotely through an in-car network; however, in most domains, manufacturers retain near-exclusive access to in-car data and require authorization (i.e., affiliation with the manufacturer) to gain more meaningful access. Even if third parties can find alternative routes to accessing car data, only a small portion of the data generated by cars adheres to open standards. The rest follows proprietary standards and requires extensive reverse engineering and ad-hoc solutions to make usable.[29] Manufacturers justify these precautions as necessary to mitigate the risk of cyber attacks, where the worst-case scenario is a crash.[30] Manufacturers negotiate individual contracts with different providers interested in using their data. The standards they promote allow the manufacturer to retain tight control over the data produced by the car.[31] Independent service providers have responded with their own secure standards for telematic car data.[32] Recent right to repair statutes like Massachusetts' Vehicle Data Access Requirement will likely force manufacturers to adopt more open standards.[33]

---

23 Spencer Chase, *Farmers Want Control of Ag Data, Survey Shows*, AgriPulse (May 11, 2016), https://www.agri-pulse.com/articles/6969-farmers-want-control-of-ag-data-survey-shows

24 Ag Data Transparent, https://www.agdatatransparent.com/ (accessed Mar. 13, 2021).

25 Kyle Wiens, *Worst In Show: Presented by Repair.org*, Repair.org (Jan. 15, 2021), https://www.repair.org/worstinshow

26 *See, e.g.* Geoffrey A. Fowler, *What does your car know about you? We hacked a Chevy to find out.*, Washington Post (Dec. 17, 2019), https://www.washingtonpost.com/technology/2019/12/17/what-does-your-car-know-about-you-we-hacked-chevy-find-out/

27 *E.g.* MetroMile, Automatic.

28 *E.g.* Connect.ai.

29 *See* FIGIEFA Automotive Aftermarket Distributors, *Commission Communication on "Free Flow of Data": Input from the Independent Automotive Aftermarket*, 8-9 (Brussels: Dec. 2016), https://www.figiefa.eu/wp-content/uploads/Free-Flow-of-Data-FIGIEFA-Input-2016_12_23.pdfFIGIEFA paper

30 Joost Vantomme & Marc Greven, *ACEA Position Paper: Access to vehicle data for third-party services", European Automobile Manufacturers Association* (Dec. 2016), https://www.acea.be/uploads/publications/ACEA_Position_Paper_Access_to_vehicle_data_for_third-party_services.pdf

31 *Id.*

32 *E.g.* Camille Sheehan, Auto Care Association and Partners to Unveil Secure Vehicle Interface at AAPEX 2018, Auto Care Association (Nov. 19, 2018), https://www.autocare.org/news/latest-news/details/2018/11/19/Auto-Care-Association-and-Partners-to-Unveil-Secure-Vehicle-Interface-at-AAPEX-2018-4693

33 Ballotpedia, *Massachusetts Question 1, "Right to Repair Law" Vehicle Data Access Requirement Initiative* (2020), https://ballotpedia.org/Massachusetts_Question_1,_"Right_to_Repair_Law"_Vehicle_Data_Access_Requirement_Initiative_(2020)

MEDICAL DEVICES

Medical devices can collect sensory data directly from a patient's body and communicate that data to feedback devices, which adjust the levels of a health intervention, or to remote data services, which allow patients and caregivers to track their health status. Participants in this workshop had particular expertise in continuous glucose monitors (CGM), wearable devices that provide real-time glucose readings for diabetics and can communicate with devices that control the dispensing of insulin.

Patients are interested in having granular access to CGM data to better monitor their own health outcomes, but device manufacturers limit the data they make available to patients and how. To address these shortcomings, open source projects like Nightscout let patients use their phones to listen for data emitted from their CGMs and send that data to the cloud for their own monitoring.[34] These open source projects use what are technically security vulnerabilities to access the data because CGMs do not offer permissioned pathways for third parties, instead only networking with devices made by the same manufacturer.[35] CGM manufacturers cite security concerns and stringent FDA requirements for not making their data available to others.[36] CGM security is regularly brought up as a source of public concern and exploits have been demonstrated at multiple hacker expos.[37]

34 Nightscout, *Welcome to Nightscout #WeAreNotWaiting*, http://www.nightscout.info (accessed Mar. 13, 2021).

35 *See* Greg Brown, *The State of Diabete Device Cybersecurity in 2019*, Healthline (Feb. 12, 2019), https://www.healthline.com/diabetesmine/cybersecurity-diabetes-devices-2019

36 *Id.*

37 U.S. Food & Drug Administration, *FDA warns patients and health care providers about potential cybersecurity concerns with certain Medtronic insulin pumps* (Jun. 27, 2019) https://www.fda.gov/news-events/press-announcements/fda-warns-patients-and-health-care-providers-about-potential-cybersecurity-concerns-certain

III. Common Challenges

Workshop participants have faced a wide range of barriers erected by data holders to make it more difficult for consumers to move data over to their platforms. As participants discussed and compared these barriers at length, a few patterns emerged. First, they noted that data holders often make data access and interoperability intentionally cumbersome for permissioned third parties but convenient for themselves and partners. Second, data holders avoid adopting open standards and sharing documentation, practices that would make it easier for third parties that do receive data to use it. Third, when data holders weaken sanctioned paths for data sharing, data users turn to unsanctioned paths, such as screenscraping, which are expensive to maintain and pose risks to themselves, consumers, and data holders alike.

Self-Preferencing

Data holders across sectors make more data available more quickly and more easily to themselves and their partners than consumer-permissioned third parties. This self-preferencing makes it difficult for consumers to try out data users' products and easy for data holders to eventually replace successful third-party services with their own homegrown alternatives. In turn, data users do not trust data holders to impartially distinguish good actors, who are safe to allow consumers to bring their data to, from bad actors.

One strategy data holders use to prevent consumers from bringing their data elsewhere is to make the process they have to go through to move their data cumbersome and unappealing. For example, health and energy data users criticized EHRs' and utilities' use of "scare screens," warnings that consumers had to click through before transferring data that used symbols and strong language to play on their fear. One energy data user noted that utilities do not show such aggressive permission screens for their own services, even when they request access to the same data for similar purposes.

Data holders can also self-preference in how quickly they make data available. This is of particular concern for businesses that interoperate with IoT devices. For many hardware applications, including tractor attachments that need to connect to the machine's controls, "live" data sent with a fraction of a second delay is useless. Data holders will often reserve more direct access to hardware for themselves, usually citing security concerns, and give data users more indirect access, often through a network. Connected car, medical device, and agricultural equipment workshop participants all raised concerns about latency in accessing data.

Data holders often extend preferential treatment to partners, making it burdensome for consumers to bring their data outside the company-sanctioned ecosystem. A participant from agriculture highlighted this dynamic in an anecdote he told about a farmer who bought seed from a vendor that provided an application that gave a "prescription" for where the seeds should be planted. The app sent the prescription data in a proprietary format that could only be read by certain brands of tractors. For these tractors, it took only fifteen minutes to transfer the data, after which the tractor's internal computer could automatically steer and plant seeds. For non-compatible tractors, though, the farmer had to spend hours manually entering data from the app.

Insufficient and Under-Adopted Standards

For data users to effectively make use of shared data, they have to know what data they will receive and how it will be formatted. If a data user wants to build a product that interoperates with multiple, similar data holders (e.g., a fintech application that works with multiple banks), it is far easier for them if the data holders' systems adhere to the same, publicly available standard. Yet data holders have tended not to adopt open standards, instead creating proprietary standards that they keep secret from others. When data holders cannot capture the value of combining multiple data sources themselves, they avoid adopting open standards in order to prevent others from doing so.[38] Participants from all six sectors criticized data holders' unwillingness to adopt useful open standards or at least document their proprietary ones.

In industries where regulators do not require data holders to follow any sort of data accessibility or interoperability standards, data holders often erect additional technical barriers to data access under the justification of security. For example, according to one participant, the big three CGM manufacturers go above and beyond federal cybersecurity requirements to intentionally design products that are difficult to reverse engineer and unable to be made "sensor agnostic." Similarly, John Deere makes it difficult for third-party attachments to interoperate with its tractors by encrypting the data it sends over its wires. As one participant noted, this is not data being sent over a network, where encryption could prevent outsiders from snooping — it only stops data from being sent to non-John Deere attachments that a farmer has explicitly chosen to plug into their tractor.

In other sectors, such as finance, agriculture, and automotive, regulators have strongly encouraged data holders to come up with standards themselves in order to improve sectoral alignment and innovation. When left to their own devices, though, data holders often come up with standards that serve themselves, not consumers or data users. In Europe, for example, automotive manufacturers came up with a connected car standard that followed the "extended vehicle" concept, which gave manufacturers widespread control over who can access the data a car produces and how that data is used.[39] American banks, which unlike their European and UK counterparts have been left to self-regulate how they make data accessible, have similarly chosen to retain near-exclusive control over their data.[40]

---

38 *See generally* Michal Gal & Daniel Rubinfeld, *Data Standardization*, 94 N.Y.U. L. Rev. 737 (2019).

39 Wolfgang Kerber, *Data Sharing in IoT Ecosystems from a Competition Law Perspective: The Example of Connected Cars* (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3445422

40 Chris Pike, *Competition and Open API Standards in Banking*, Organization for Economic Co-Operation and Development (OECD) - Competition Division, 7-8 (Mar. 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3487628

Even when a standard is legally required by regulators or is industry ubiquitous, data holders can undercut its effectiveness through poor upkeep and implementation. In the energy industry, this happened with Green Button, a standard launched in 2012 that among other things let consumers access their electricity usage data digitally and bring it to third-party services. When Green Button was established almost a decade ago, the uses of energy data were only theoretical. Today, practical experience has changed considerations about what data is useful and how quickly it should be made available, yet utilities have shown little interest in updating the standard. A similar dynamic is playing out in agriculture — modern tractor attachments are pushing the limit of the ubiquitous ISOBUS standard, which allows tractors to communicate with tows, but John Deere and other tractor companies are unwilling to update to the latest version of the standard or to extend it to front-facing tractor attachments.

Healthcare regulators have taken a different approach from other sectors by empowering a standards body to maintain EHR interoperability standards and keep them up to date. Even though participants see the body as uncaptured by data holder interests, EHR providers can still drag their feet with poor implementation in what one workshop participant called "death by a thousand cuts." Energy participants said they face the same issue with utilities, which sometimes so underinvest in data sharing systems that the data they send is incorrect, incomplete, or months late.

Unsanctioned Data Collection

When data holders do not give consumers an easy way to share their data with third parties, those third parties use their own methods to access data, such as screenscraping and reverse engineering. Consumers have to give data users permission for these methods to work, but they still raise two salient problems. First, consumers do not always understand the extra security and privacy risks these methods entail. Second, unsanctioned pathways are expensive for data users to maintain, and data holders can selectively make them more or less difficult for certain firms in potentially anti-competitive ways.

Screenscraping is a particularly common and problematic practice in the financial tech sector. Fintech companies that screenscrape banking data often have completely unfettered access to a consumer's account, allowing them to collect data with impunity. With so much data and so little regulation controlling how they use it, data users can become powerful data aggregators, and according to workshop participants, use data in ways that consumers may not have anticipated or would be happy with.

There are legal and technical means to prevent screenscraping, and data holders can threaten competition through selective enforcement of such policies. In the energy sector, it is an open secret that many third-party applications access energy usage data through screenscraping, in part due to the insufficiencies and poor adoption of Green Button. Utilities often turn a blind eye to commercial and industrial customers who screenscrape but prevent residential customers from doing the same because they know it likely comes from third-party applications encroaching on their potential future market share.

Reverse engineering raises its own concerns. For example, medical device data users sometimes build products that depend on what are technically device security vulnerabilities. In the case of CGMs, this means that consumers who want to use certain third-party open source monitoring software have to use unpatched devices that are years out of date.[41] There is enough demand for this software that it has created a black/gray market for these outdated devices.

Besides raising security concerns, reverse engineering can also be expensive for data users, and in many industries, the ability to send over-the-air updates has created an expensive arms race between data holders and data users. According to one agricultural equipment developer in attendance, it costs between $800,000 and $1 million dollars to reverse engineer one type of tractor to make it interoperable with one header. And today, when original equipment manufacturers update their tractors remotely, all of that work can be undone with one download.

41 *E.g.* OpenAPS, https://openaps.org/ (accessed Mar. 13, 2021).

IV. Common Policy Concerns

In addition to facing common challenges, participants had common concerns about trends in data sharing policymaking. First, participants' experiences led them to believe that the market does not naturally encourage better data sharing practices, even when there is robust competition between data holders. Second, participants agreed that the question of data ownership has distracted from the more important question of data access. Third, participants criticized regulators for siloing conversations about data sharing and competition from conversations about privacy and cybersecurity.

Competition between data holders has not improved data access

Multiple workshop participants disagreed with a common argument proffered by data holders — that if sharing data with third parties benefits consumers, the market will encourage it naturally. Participants found that in practice, more competition among data holders has not made for better data sharing practices, despite the fact that interoperability with ancillary services should make for a better consumer experience. Whether data holders are regulated monopolies (energy companies), dominated by a few players (EHR vendors, medical device manufacturers, agricultural equipment manufacturers), or have dozens or thousands of competitors (car companies, banks), they are incentivized to be conservative in how they share data.

Participants suggested a few reasons why this might be the case. Some believed that data holders withhold data because they do not consider data sharing worth the potential liability. One fintech participant argued that this was particularly true in the case of banks — if a data user has a breach, studies show that consumers want to be able to sue the deep-pocketed data holder (i.e., the bank), not the cash-poor data user.[42]

42 The Clearing House, *Fintech Apps and Data Privacy: New Insights from Consumer Research*, 15 (Aug. 2018), https://www.theclearinghouse.org/-/media/New/TCH/Documents/Data-Privacy/TCH-Consumer-Research-Report-08-20-2018.pdf

Other participants suggested that data holders want to retain exclusive control of the innovative potential of data for themselves. As one participant explained in the case of energy, utility companies "are expanding their monopoly on a de facto basis" by using data to reach into new domains, including solar. Agriculture and automotive data holders also claim that exclusive control of data incentivizes them to use it to create new value for customers that only they can create. One agriculture participant pointed out that these new value propositions almost never come to fruition, noting how agricultural equipment companies have claimed for years that they will offer part prediction failure but never have.

The only advantage of robust competition between data holders that participants mentioned was that it makes it more difficult for them to collude to not share data. One researcher in attendance explained how this happened in the case of automobile dealership management systems, where the two dominant firms were able to agree not to make their data portable.[43]

Data ownership debates are a distraction

One point of contention between the researchers and the practitioners in attendance was about whether or not policymakers should focus on whether consumers "own" their data. Practitioners generally agreed that it was the wrong question to ask. As one workshop participant put it, "Once you ask the question of who owns the data, you validate the idea that data can be owned…we need to stop this paradigm and just ask who can access the data." Another participant said that in healthcare, state laws that give patients ownership over their data have had little effect on how easy it is in practice for patients to send their data to third parties. A third participant argued that data ownership doesn't address questions of how data is accessed and overlooks the issues of standards and timeliness.

The practitioners agreed that they want independent, unfettered access to any data a consumer wants to bring over, but there was disagreement about how to achieve that. In the software session, there was debate over the effectiveness and enforceability of fair, reasonable, and non-discriminatory access requirements. In the IoT session, attendees disagreed about whether the law should require data holders to make their systems interoperable or whether third parties should be given more legal leeway to reverse engineer.

Data sharing is siloed from privacy and cybersecurity

Participants noted that in policymaking circles, conversations about data sharing occur separately from conversations about privacy and security. When data users complain about overly restrictive data sharing practices, data holders gesture to larger conversations with regulators about security and privacy that the data users were not involved in. At best, this creates data policy that undoes with one hand what it does with the other. At worst, it creates opportunities for bad faith information blocking in the name of privacy and security.

Attendees pointed to multiple examples of data holders employing this "you wouldn't understand" line of argument. EHRs, for example, have made misleading claims about how data users are subject to HIPAA. In agriculture, John Deere wields the flag of privacy to claim that it protects farmers from dangerous third-party data aggregators. Yet in fact, as one workshop participant noted, John Deere is simply reserving the role of aggregator for itself.

---

43 Peter Swire, *The Portability and Other Required Transfers Impact Assessment: Assessing Competition, Privacy, Cybersecurity, and Other Considerations*, Georgia Tech Scheller College of Business Research Paper No. 3689171 40-41, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3689171

V. Looking Ahead

Again and again, workshop participants found that their data sharing problems are more similar than they are different. The Data Sharing Practitioners Workshop revealed that despite coming from different sectors, data holders employ the same few strategies and play out the same few dynamics with data users. Data sharing policy is enormously complex, but this paper should give a glimmer of hope that what works in one sector may work in another.

Though the goal of this workshop was to be descriptive rather than prescriptive, it is worth highlighting a few regulatory interventions that participants expressed interest in. This is not meant to be an endorsement of any one path forward — rather it is meant to shed light on the priorities of data users.

1. *Empower existing, uncaptured standards bodies*. Workshop participants applauded the 21st Century Cures Act for putting the weight of law behind mature EHR data sharing standards (e.g., SMART on FHIR) and the organizations that maintain them. Energy sector participants noted how the case of Green Button shows that when the government empowers a standard without a body to maintain it, it can quickly go out of date. At the same time, the Cures Act avoided the kind of broad interoperability mandate that IoT participants fear data holders could easily circumvent.
2. *Allow for alternative data access methods*. Workshop participants wanted new ways to split the difference between the unfettered data access of unsanctioned data pathways, such as screenscraping and reverse engineering, and the security of sanctioned pathways, such as APIs and open hardware interfaces. Participants expressed interest in middle ground alternatives like delegated account access and right to repair laws.
3. *Improve data sharing implementation*. Workshop participants were not just interested in what data is accessible but also how easy it is for engineers to use and for consumers to give third parties access to. IoT participants wanted data holders to release their documentation. In sectors like healthcare and energy that already have some data sharing requirements, data users were interested in improving the consumer experience of permitting data transfer.

As American regulators and lawmakers have conversations about what the next generation of data policy looks like, it is critical that they bring in the perspectives of firms that use shared data, not just the firms that hold it. The Data Sharing Practitioners Workshop showed just how much innovation is already occurring with data that consumers share, and just how worthwhile it is to consider the opinions of data users are on how to grow that innovation further. This workshop paper has begun the work of coalescing the viewpoints and values of data users from different sectors. It is up to policymakers to invite data users to keep giving their input into the discourse.

Engelberg Center on Innovation Law & Policy
139 MacDougal Street Room 408 | New York, NY 10012

NYUEngelberg.org

engelberg.center@nyu.edu

@NYUEngelberg

ENGELBERG CENTER
on Innovation Law & Policy
NYU School of Law