Testimony

Hearing on Data Ownership: Exploring Implications for Data Privacy Rights and Data Valuation

WILL RINEHART | OCTOBER 24, 2019

Testimony to the Committee on Banking, Housing, and Urban Affairs

United States Senate

* *The views expressed here are my own and do not represent the position of the American Action Forum.*

Chairman Crapo, Ranking Member Brown, and Members of the Committee, thank you for the opportunity to testify today regarding data property rights. Like many privacy experts, I'm skeptical that data property rights are the best policy mechanism for ensuring privacy is secured in the digital age. I hope to make three main points today:

- A property right to personal data isn't needed to establish consumer privacy rights, nor would it be economically efficient to establish this kind of property right;
- Valuing personal data is difficult because raw or personal data per se is not what is in demand, but rather the insights that can be gleaned from that data—insights that often depend on the data's environment; and
- Regardless of the particular policy mechanism, privacy laws will create unavoidable costs from compliance, which will impact investment opportunities in countless industries.

The Purposes and Limitations of Propertization

With Congress again considering federal privacy legislation, the idea of personal data property rights is being explored as one policy mechanism for securing privacy.[1] The very phrase "personal data" conjures up the notion that individuals own that data and firms are merely taking it. Data propertization, which is the creation of property rights in law, has been seen as an attractive alternative since the 1970s, for two reasons.[2] First, it would grant individuals the ability to sell their personal data, thus allowing them to recapture some of its value. Second, propertization would force companies to internalize the costs of disclosure, thereby aligning firm and user expectations about data collection and use since users would be able to bargain over the terms of the deal.[3]

There are reasons to be skeptical that assigning property rights in data will have unalloyed benefits. For one, assigning property rights to data is a contortion of the normal reasoning that underpins intellectual property (IP) rights such as copyright and patents. Information, which is embodied in copyrights and patents as well as user data, can be easily reproduced (i.e. is non-rivalrous), and it is difficult to prevent non-paying consumers from accessing it (non-excludable). Property rights incentivize information creation, since those rights give the holder the ability legally to exclude others, thus making the information rivalrous. Yet, the problem faced in privacy is of the opposite kind—the purpose is to limit information disclosure.

Second, it is unclear if the assignment of data property rights will align incentives between users and firms. While it is the case that information disclosure can either be beneficial or detrimental, users cannot know beforehand if the use of their data will necessarily lead to better products.[4] Data propertization would only exacerbate this problem, forcing users to search for the best value for their data. In other words, data property rights would make users data entrepreneurs. Searching for innovative opportunities is costly, and thus one could imagine that users will likely hire an intermediary to do this task—which is the job of platforms and other data providers presently.

As is detailed in an appendix to this paper, the Hart-Grossman-Moore model of property helps to flesh out the idea. This model can help to determine where it is most efficient to allocate property rights. When one party's investment in the data does not boost the total value that much, then it is better for the other party to have control of the assets. In the parlance of economics, the party with higher marginal returns from investment should be given the rights of control, which is why platforms, and not users, spend so much time and effort to understand what is happening on the platform. Assigning data property rights to users will likely be inefficient because it will change the investment decision veto point.

Third, and most important for this Committee, real world implementation will prove tricky because of the interconnected nature of information. The vast majority of data generated in the last decade comes from user interactions with online platforms. If Google didn't exist, there would be no search data. If Facebook didn't exist, there wouldn't be social graph data. To understand the challenge of implementing data property rights, it is helpful to recognize how three classes of data interact in online platforms.[5] *Volunteered data* is data that is both innate to an individual's profile, such as age and gender, and information they share, such as pictures, videos, news articles, and commentary. *Observed data* comes as a result of user interactions with the volunteered data; it is this class of data that platforms tend to collect in data centers. Last, *inferred data* is the information that comes from analysis of the first two classes, which explains how groups of individuals are interacting with different sets of digital objects. At the very least, then, data is a co-created asset, with the users providing volunteered data and the platform assembling observed data to create inferred data. Creating data property rights will likely necessitate that only one party has rights, which has been a sticking point for previous efforts.

As the German government discovered when trying to implement data property for connected cars, determining the owner isn't simple. Does the car company own the property right, or might it be the driver, or even the rider? [6] Privacy scholar Robert Gellman demonstrated this problem would also beleaguer health care. For example, information about a child's health could simultaneously belong to the patient, the patient's family, the school, the pharmacy, the supermarket, the pediatrician, the drug manufacturer, social media platforms, advertising companies, or Internet service providers.[7] Questions of fuzzy ownership continually plague IP and would similarly afflict data property.

Further, and even more practically, a property right in data isn't needed to establish consumer privacy rights. For evidence of this fact, one only needs only to consult the current laws in the United States. The Gramm-Leach-Bliley Act (GLBA), the Fair Credit Reporting Act (FCRA), the Children's Online Privacy Protection Act (COPPA), and the California Consumer Privacy Act (CCPA), just to name a few, all protect privacy without creating property rights. As Stanford Law Professor Lothar Determann has said quite bluntly, "no one owns data" because data are already "subject to a complex landscape of access rights and restrictions."[8] Privacy regulation already defines certain kinds of entitlements to control and contract upon data. Adding a superordinate property right on top of these existing restrictions would make the entire enterprise all that more complicated and undermine current efforts to grant consumers control. If data property rights were implemented, for example, would an individual be able to limit critical information from being shared with

credit rating agencies?

Determann isn't the only scholar of privacy who opposes propertization efforts. Technologist Larry Downes has been critical of the idea and instead prefers the current licensing model since it "recognizes that most information with economic value is the collaborative creation of multiple sources, including individuals and service providers."[9] Law Professor Julie Cohen has argued against privacy as property as well since it doesn't uphold the values of autonomy and participation that are so central to privacy.[10] European law professor Bart Schermer agreed with Cohen when the issue was raised in 2015 as an alternative to the European Union's (EU) General Data Protection Regulation (GDPR), saying that "Reducing the discussion about privacy and personal data to a discussion about ownership oversimplifies the discussion about privacy in the information society and may lead to sub-optimal results when it comes to regulating the use of personal data."[11] But Dr. Mark MacCarthy of Georgetown University said it best, laying out the world of data property rights as "a privacy nightmare rather than a privacy paradise."[12] As much as there is disagreement in privacy advocacy and scholarship, there is consistent agreement that propertizing data has serious limits.

Finally, pricing data, which is one stated goal of data property rights, will have deleterious effects on privacy expectations. As Jason Aaron Gabisch and George R. Milne reported in the Journal of Consumer Marketing, "The findings show that receiving compensation, especially when it is a monetary reward, reduces consumer expectations for privacy protection."[13]

Valuing Data

Four methods can be employed to value intangibles such as data: income-based methods, market rates, cost methods, and shadow prices.

Most popular data valuations are accomplished through income derivations, often by simply dividing the total market capitalization or revenue of a firm by the total number of users. For those in finance, this method seems most logical since it is akin to an estimate of future cash flows. In a *Wired* article, for example, Antonio Garcia Martinez placed an upper bound of $112 on the value of data for users in the United States, citing Facebook's 2018 annual report.[14] Similarly, when Microsoft bought LinkedIn, reports suggested that it was buying monthly active users at a rate of $260 per user.[15] Stanford Law Professor A. Douglas Melamed argued before the Senate Judiciary that the upper-bound value on data should at least be cognizant of the acquisition cost for advertisements—putting the total value per user at around $16.[16]

Still, these income-based valuations aren't exact estimates because they are not capturing a user's ability to marginally earn revenue, which is where the price would be set. As noted before, inferential data is the key for platform operators, as it drives advertising decisions and helps determine what content is presented to users. Thus, the ultimate value of a user's data would combine the value of that user's data to increase all their friend's demand for content and the value of that user's data to contribute to increases in advertising demand. Calculating marginal income valuations in this manner are difficult, but Shapley values have been shown as a viable method theoretically.[17],[18] Still, it remains unclear if firms would be able to implement this method on their platform.[19] Needless to say, income-based valuations are difficult.

Second, market prices are another method of valuing data, and they tend to place the lowest premium on data. For example:

- Vice recently reported that Departments of Motor Vehicles across the United States have been selling

individual records for as little as one cent each;[20]

- *Wired* editor Gregory Barber sold his location data, Apple Health data, and Facebook data, and all he got was a paltry $0.003 for everything together;[21]
- After a breach at Facebook, Facebook logins were selling on the dark web for $2.60 per user;[22]
- Advertisers typically pay $0.005 for complete profile for an individual;[23]
- General information about a person, such as their age, gender, and location is worth a mere $0.0005 per person, or $0.50 per 1,000 people;[24]
- Auto buyers are worth about $0.0021 per person, or $2.11 for every 1,000 people;[25] and
- For $0.26 per person, buyers can access lists of people with specific health conditions or taking certain prescriptions.[26]

In reviewing these estimates, *The Financial Times* noted that "the sum total for most individuals often is less than a dollar." It is worth noting that sub $1 payments have been unprofitable for firms to process due to the fixed technical costs for developing the backend architecture and hardware, storage costs for transaction integrity and legal purposes, computational costs for processing payments, communication costs for information transfer, and administrative costs.[27]

As with any market, it is important to pay attention to the difference between the clearing price and the asking price. The bankruptcy proceedings for Caesars Entertainment, a subsidiary of the larger casino company, offers a unique example of this problem. As the assets were being priced in the selloff, the Total Rewards customer loyalty program got valued at nearly $1 billion, making it "the most valuable asset in the bitter bankruptcy feud at Caesars Entertainment Corp."[28] But the ombudsman's report understood that it would be a tough sell because of the difficulties in incorporating it into another company's loyalty program. Although it was Caesar's asset with the highest valuation, its real value to an outside party was an open question.

The Total Rewards example underscores an important characteristic of data: It is often valued within a relationship but is difficult to value outside of it. Within economics, there is a term for this phenomenon, as economist Benjamin Klein explained: "Specific assets are assets that have a significantly higher value within a particular transacting relationship than outside the relationship."[29] Asset specificity helps to explain why there isn't an auction market for personal data. It isn't the raw data that is in demand, but the insights that can be gleaned from that data.

Third, data might be valued using cost-based methods, but this method also has shortcomings. Proxying the value of data by summing the salaries of data analysts and the costs of data centers will likely underestimate the value of data. Data is an intermediate product for other business processes. In practice, cost-based methods would probably look like Shapley values anyway.

Last, data can be valued through shadow prices.[30] For those items that are rarely exchanged in a market, prices are often difficult to calculate, so other methods are used to appraise what is known as the shadow price. For example, a lake's value might be determined by the total amount of time in lost wages and money spent by recreational users to get there. Similarly, the value of social media data might be calculated by tallying all of the forgone wages in using the site. A conservative estimate from 2016 suggests that users spend about fifty minutes a day month on Facebook properties.[31] Since the current average wage is about $28, this calculation indicates that people roughly value the site by about $8,516 over the entire year.[32] A study using data from 2016 using similar methods found that American adults consumed 437 billion hours of content on ad-supported media, worth at least $7.1 trillion in terms of foregone wages.[33]

Shadow prices can also be calculated through surveys, which is where this method gets particularly controversial. Depending on how the question is worded, users' willingness to pay for privacy can be wildly variable. Trade association NetChoice worked with Zogby Analytics to find that only 16 percent of people are willing to pay *any* price for online platform service.[34] Strahilevitz and Kugler found that 65 percent of email users, even though they knew their email service scans emails to serve ads, wouldn't pay for an alternative.[35] As one seminal study noted, "most subjects happily accepted to sell their personal information even for just 25 cents."[36] Using differentiated smartphone apps, economists were able to estimate that consumers were willing to pay a one-time fee of $2.28 to conceal their browser history, $4.05 to conceal their list of contacts, $1.19 to conceal their location, $1.75 to conceal their phone's identification number, and $3.58 to conceal the contents of their text messages. The average consumer was also willing to pay $2.12 to eliminate advertising.[37]

In all, there is no one single way to estimate the value of data, and none of them is particularly easy to implement.

The Impact of New Privacy Laws

Regardless of the path that is taken, new privacy laws will have both direct and indirect impacts on the economy, best seen in the wake of the GDPR and estimates from the CCPA. First, privacy regulation will force firms to retool data processes, known as refactoring, to comply with new demands. This refactoring is generally a one-time fixed cost that raises the cost of all information-using entities. Second, the regime will add risk compliance costs, causing companies to staff up to ensure compliance. Finally, privacy laws change the investment dynamics of the affected industries, as the market shifts to account for the newly expected returns.

Currently, the retooling costs and risk compliance costs are going hand in hand, so it is difficult to determine the costs of each. Still, they are substantial. A McDermott-Ponemon survey on GDPR preparedness found that almost two-thirds of all companies say the regulation will "significantly change" their informational workflows. According to this survey, the average budget for getting to compliance tops $13 million. The International Association of Privacy Professionals estimated that GDPR will cost Fortune 500 companies around $7.8 billion, and these won't be one-time costs since "Global 500 companies will be hiring on average five full-time privacy employees and filling five other roles with staff members handling compliance rules." A PwC survey on the rule change in Europe found that 88 percent of companies surveyed spent more than $1 million on GDPR preparations, and 40 percent more than $10 million.

Refactoring and compliance costs are adding up for CCPA as well. California's standardized regulatory impact assessment (SRIA) for CCPA calculated the total costs at $55 billion, which is nearly 1.8 percent of the total gross state product.[38] The range of affected firms is massive. On the bottom end of the estimate, 15,643 businesses could feel an impact. On the top end, 570,066 companies will have to come into compliance with the law. Most alarming, the authors conclude that "economic impact of the regulations on these businesses located

outside of California [that serve California consumers] is beyond the scope of the SRIA and therefore not estimated." If something akin to the California law were applied to the United States, the Information Technology and Innovation Foundation estimated the cost at $122 billion per year.[39]

Finally, privacy laws will surely change the investment and market dynamics in countless industries. When the EU adopted the e-Privacy Directive in 2002, Goldfarb and Tucker found that advertising became far less effective, which reverberated throughout the ecosystem as venture capital investment in online news, online advertising, and cloud computing dropped by between 58 to 75 percent. In Chile, for example, credit bureaus were forced to stop reporting defaults in 2012, which was found to reduce the costs for most of the poorer defaulters, but raised the costs for non-defaulters. Overall the law led to a 3.5 percent decrease in lending and reduced aggregate welfare. Early research on the GDPR has also found drops in investment. While much smaller than the United States, EU venture funding decreased by decreased 39 percent while the total number of deals saw a 17 percent drop.[40]

Conclusion

The dilemma for this Committee and others within Congress is hardly enviable. America's privacy pandect is complex, making difficult the task of creating new laws to enhance consumer privacy. While there is much disagreement in the privacy community, there is widespread agreement that data property rights are an unwieldy way of doing things. There should be no delusions, however, about the impacts. There will be serious costs involved with any new law. As Seth Godin once remarked, "The art of good decision making is looking forward to and celebrating the tradeoffs, not pretending they don't exist." That is sage advice for any privacy legislation.

Technical Appendix

One way to understand this bargain is through the Grossman-Hart-Moore model, which considers a relationship between two risk-neutral parties, a buyer and a seller, or *B* and *S*. For this exercise, let's assume that the buyer of the data, *B*, is the platform, and the seller of the data, *S*, is the user, and again let's just work with the singular transaction. As such, the platform buys data, which is an intermediate good, from the users to create a final output. The value of the final good is *V(e)*, which is contingent on *e*, a variable for the investment into the process by the platform. Similarly, the cost of the intermediate good is *C(i)*, which is contingent on the investment, *i*, in the process conducted by the user.

There are two periods. In the first period, each party undertakes some kind of investment and in the second period, they decide to trade at a specific price, *p*. If they don't end up trading, they can turn to others and do so. A key assumption of this model is that the investments in the first time period are not contractible.

The social optimum would involve maximizing the total benefits minus the investment costs:

$$max_{e,i}\ \{V(e) - C(i) - e - i\}$$

Optimal investment thus occurs when

$$V'(e) = 1 \text{ and } -C'(i) = 1$$

But in the present set up, each party will only retain half of the gains from trade, such that

$$max_e \left\{\frac{V(e)}{2} - e\right\} \text{ and } max_i \left\{\frac{-C(i)}{2} - i\right\}$$

Because the parties will have to bargain over how to split the total surplus, each will get half of the benefits from their investment. See Aghion and Holden (2011) for further details on the Nash bargaining.[41] Thus, each party will underinvest relative to the first best.

If the parties instead have vertically integrated, the result is slightly different. If, say, *B* controls the total gains from the production processes, then *B* will invest at their first best level while *S* will underinvest. Similarly, if *S* were to own total gains, then *S* will invest at their first best, while *B* will underinvest.

This model yields some interesting insights. It is important to note that, like the rest of the literature in this space, the investment elasticities are key. Since *S* or users, have extremely inelastic investment decisions, that is, they don't change that much with the possibility of *B* appropriating them, it is the case that *B* should own the total gains.

This makes sense in the case of platforms. The investment that matters the most lies in the inference data of the platform. Users have indeed tried to sell their own "investment," but these transactions don't yield much. Moreover, the relative investments speak to why data ownership efforts are likely to fail. Since the marginal returns for any user *S* is much higher when a platform *B* controls both, as compared to when users simply "own their data," independent ownership is likely to lead to inefficient gains for all sides.

[1] Michael Gorthaus, "Andrew Yang proposes that your digital data be considered personal property," available at: https://www.fastcompany.com/90411540/andrew-yang-proposes-that-your-digital-data-be-considered-personal-property.

[2] Alan Westin, *Information Technology in a Democracy*.

[3] Pamela Samuelson, "Privacy As Intellectual Property," available at: http://people.ischool.berkeley.edu/~pam/papers/privasip_draft.pdf.

[4] Alessandro Acquisti, Curtis R. Taylor & Liad Wagman, "The Economics of Privacy," available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2580411.

[5] World Economic Forum, "Personal Data: The Emergence of a New Asset Class," available at: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

[6] Lothar Determann, "No One Owns Data," available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3123957.

[7] Robert Gellman, "Health Information Privacy Beyond HIPAA: A 2018 Environmental Scan of Major Trends and Challenges," available at: https://ncvhs.hhs.gov/wp-content/uploads/2018/05/NCVHS-Beyond-HIPAA_Report-Final-02-08-18.pdf

.

[8] See footnote 6.

[9] Larry Downes, “A Rational Response to the Privacy ‘Crisis,’” available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2200208.

[10] Julie E. Cohen, “Examined Lives: Informational Privacy and the Subject as Object,” available at: https://scholarship.law.georgetown.edu/cgi/viewcontent.cgi?article=1819&context=facpub.

[11] Bart Schermer, “Privacy and property: do you really own your personal data?” available at: https://leidenlawblog.nl/articles/privacy-and-property-do-you-really-own-your-personal-data.

[12] Mark MacCarthy, “Privacy Is Not A Property Right In Personal Information,” available at: https://www.forbes.com/sites/washingtonbytes/2018/11/02/privacy-is-not-a-property-right-in-personal-information/.

[13] Jason Aaron Gabisch & George R. Milne, “The impact of compensation on information ownership and privacy control,” available at: https://www.emerald.com/insight/content/doi/10.1108/JCM-10-2013-0737/full/html.

[14] Antonio Garcia-Martinez, “No, Data Is Not the New Oil,” available at: https://www.wired.com/story/no-data-is-not-the-new-oil/.

[15] James E. Short & Steve Todd, “What’s Your Data Worth?” available at: https://sloanreview.mit.edu/article/whats-your-data-worth/.

[16] A. Douglas Melamed, “Prepared Statement,” available at: https://www.judiciary.senate.gov/download/melamed-testimony.

[17] Amirata Ghorbani & James Zou, “Data Shapley: Equitable Valuation of Data for Machine Learning,” available at: https://arxiv.org/abs/1904.02868.

[18] Eric Bax, “Computing a Data Dividend,” available at: https://arxiv.org/pdf/1905.01805.pdf.

[19] While Bax has shown that Shapley values can be implemented in polynomial time, it is unclear if Shapley values that exhibit demand interdependencies could be implemented in polynomial time as well.

[20] Joseph Cox, “DMVs Are Selling Your Data to Private Investigators,” available at: https://www.vice.com/en_us/article/43kxzq/dmvs-selling-data-private-investigators-making-millions-of-dollars?utm_campaign=sharebutton.

[21] Gregory Barber, “I Sold My Data For Crypto, Here’s How Much I Made,” available at: https://www.wired.com/story/i-sold-my-data-for-crypto/.

[22] Dan Hall, “Hackers selling Facebook logins on the dark web for $2,” available at: https://nypost.com/2018/10/01/hackers-are-selling-facebook-logins-on-the-dark-web-for-2/.

[23] Frank Pasquale, “The Dark Market for Personal Data,” available at: https://www.nytimes.com/2014/10/17/opinion/the-dark-market-for-personal-data.html.

[24] Financial Times, “Financial worth of data comes in at under a penny a piece,” available at: https://www.ft.com/content/3cb056c6-d343-11e2-b3ff-00144feab7de.

[25] Ibid.

[26] Ibid.

[27] Ioannis Papaefstathiou, “Evaluation of Micropayment Transaction Costs,” available at: http://web.csulb.edu/journals/jecr/issues/20042/Paper3.pdf.

[28] Kate O’Keeffe, “Real Prize in Caesars Fight: Data on Players,” available at: https://www.wsj.com/articles/in-caesars-fight-data-on-players-is-real-prize-1426800166.

[29] Benjamin Klein, “Asset specificity and holdups,” available at: http://masonlec.org/site/files/2012/05/WrightBaye_klein-b-asset-specificity-and-holdups.pdf.

[30] Anthony E. Boardman, David H. Greenberg, Aidan R. Vining, & David L. Weimer, *Cost benefits Analysis Concepts and Practice*.

[31] James B. Stewart, “Facebook Has 50 Minutes of Your Time Each Day. It Wants More.” available at: https://www.nytimes.com/2016/05/06/business/facebook-bends-the-rules-of-audience-engagement-to-its-advantage.html.

[32] Bureau of Labor Statistics, “Average hourly and weekly earnings of all employees on private nonfarm payrolls by industry sector, seasonally adjusted,” available at: https://www.bls.gov/news.release/empsit.t19.htm.

[33] David S. Evans, “The Economics of Attention Markets,” available at: https://www.competitionpolicyinternational.com/the-economics-of-attention-markets/.

[34] NetChoice, “American Consumers Reject Backlash Against Tech,” available at: https://netchoice.org/american-consumers-reject-backlash-against-tech/.

[35] Lior Stahilevitz & Matthew B. Kugler, “Is Privacy Policy Language Irrelevant to Consumers?” available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2838449.

[36] Jens Grossklags & Alessandro Acquisti, “When 25 Cents is too much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information,” available at: https://www.econinfosec.org/archive/weis2007/papers/66.pdf.

[37] Scott J. Savage & Donald M Waldman, “The Value of Online Privacy,” available at: https://static1.squarespace.com/static/571681753c44d835a440c8b5/t/5735f456b654f9749a4afd62/1463153751356/The_

.

[38] Berkeley Economic Advising and Research, “Standardized Regulatory Impact Assessment: California Consumer Privacy Act of 2018 Regulations,” available at: http://www.dof.ca.gov/Forecasting/Economics/Major_Regulations/Major_Regulations_Table/documents/CCPA_Regula SRIA-DOF.pdf.

[39] Alan McQuinn & Daniel Castro, “The Costs of an Unnecessarily Stringent Federal Data Privacy Law,” available at: https://itif.org/sites/default/files/2019-cost-data-privacy-law.pdf.

[40] Jian Jia, Ginger Jin & Liad Wagman, “The short-run effects of GDPR on technology venture investment,” available at: https://voxeu.org/article/short-run-effects-gdpr-technology-venture-investment.

[41] Philippe Aghion & Richard Holden, “Incomplete Contracts and the Theory of the Firm: What Have We Learned over the Past 25 Years?” available at: https://www.aeaweb.org/articles?id=10.1257/jep.25.2.181.