A New Way Forward: The Internet & Data Economy

by
Alexa Reese Canaan

Attestation:

Summary:

In this paper, I argue for an existing technology that could be used in novel ways. Specifically, I argue for a restructuring of the internet and the data economy that is foundational to the internet. I believe that because data is valuable and providing data is labor, data citizens (general internet users) who provide data to technocrats (large internet platforms) should be explicitly compensated for their data by technocrats, rather than accepting the current software-as-a-service model where data citizens use platforms for free, but have to (sometimes unknowingly) relinquish large amounts of data to technocrats that they rely on to make the platforms function. Currently, we are in one of the worst positions, where both platforms and data citizens are losers in the market because it is not running efficiently. Making changes to the system benefits everyone, even if there is heavy initial investment into changing this system. I first give background on how this system came to be, without much organization at all and illustrate that these large platforms have become indispensable to our lives, and therefore the protections that extend to these platforms should have particular nuance. There are legal, economic, and political reasons for why this shift makes sense as well. Next, I evaluate the lump sum payment, pay-per-data point, and dividend models and how data citizens and technocrats are affected by these different structural changes. Finally, I conclud that the dividend model is the most feasible and easiest to implement for both data citizens and technocrats.

A New Way Forward: The Internet & Data Economy

<u>Intro:</u>

The emergence of "big data" has rapidly changed the way our world operates. 289,351 gigabytes of data are created on the internet every second (Department of Economic and Social Affairs 2019). An entire economy has risen up around the emergence of data. Big platforms on the internet use "big data" and turn it into value through a mechanism called software-as-a-service, where consumers give their data to these companies in exchange for some service. This structure has made people more reliant on the internet, making it almost impossible for consumers to opt out of using services provided by technocrats like Facebook and Google, especially when those companies are conglomerates of smaller alternative companies that they've bought up (Posner and Weyl 2018). Data on the internet is currently treated as a public good, where someone's use of it does not prevent someone else from using it and it can be used over and over again. General internet users (a.k.a data citizens) cannot prevent technocrats or other data citizens from using it repeatedly. Without intervention, public goods are exploited (Department of Economic and Social Affairs 2019). Therefore, we need a new structure for the data economy where data citizens and technocrats can both "win".

I propose a new way forward for the data economy where data citizens are explicitly paid by large platforms for the use of their data and have transparency into what their data is being used for. Since data is valuable and providing value is labor, data citizens should be able to decide if and when they want to provide their data, decide which aspects of their data to give, opt out any time, and have some leverage in terms of deciding how they are compensated for this data. The current system gives no bargaining power to data citizens. For economic, political, and legal reasons, we as a society must establish an explicit data economy between data citizens and technocrats where consumers own their own data and are compensated monetarily. While this may be considered tedious or more work in the short run, creating a more equitable and transparently transactional data economy will distribute earnings from this economy towards more of the users who directly contribute to it and make it available to those who cannot currently afford to participate in this system.

<u>Background:</u>

Our current system falls under an idea posed by Posner and Weyl as "technofeudalism". They argue that platforms give users just enough free service to keep them using said service, ensuring dependency. This system has come to be due to the history of how the internet started. A key principle of the internet has always been low barriers to participation. Therefore, it was hard for the first websites to encourage people to pay for services, or otherwise block them out. Plus, in the beginning, internet service was small and payments were costly and therefore not justifiable. With the added threat of hackers, transacting online didn't make sense. Eventually, companies ended up turning to ad revenue. The result was that companies evolved to maximize profits from ad revenue, giving users just enough free service to stay active on the platform. Games and relaxing structures of websites arose as a way for companies to get data from customers more efficiently. Websites aren't incentivized to make their platforms better for customers until it impairs the user experience so much that users stop accessing these platforms.

Motivations:

There are several existing tenants in the legal, economic, and political realms that show favorability towards changing the current system.

Legal

The current structure of the data economy has been debated with court cases that support and do not support changing the structure of the data economy, and other legal precedents that require more research. There are two important pieces of European Union (EU) legislation that have pushed technocrats to give up some of their power: The European Data Governance Act and the General Data Protection Regulation (GDPR). The European Data Governance Act benefits data citizens by creating a licensing regime for data intermediaries, organizations that organize commercial arrangements between data holders and data users. Not only that, but the Act also grants data intermediaries the ability to determine if technocrats offer appropriate protections for non-personal data and can restrict access to non-personal data until there are adequate protections (European Commission 2022). The EU is already forcing technocrats to check themselves before needlessly accessing data on EU citizens. The GDPR adds more guidance, instructing that companies collecting data should collect and process only as much data as

necessary for the purposes specified and stored for only as long as necessary (Wolford 2018). It requires that users knowingly consent to giving up their data to these companies as well. This law is the first step in holding companies accountable for data protection, granting data citizens "the right to be informed, the right of access, the right to rectification, the right to erasure, the right to restrict processing, the right to data portability, the right to object and also rights around automated decision making and profiling" (Burgess 2020). These guidelines require more transparency and accountability from platforms, however under these mechanisms data citizens may have to sacrifice using a service altogether or giving away their data for free, although they know what would happen to their data. These policies are the beginning frameworks to continue to build off of on a global scale.

A legal case that proved promising for being taken up in the first place, but ultimately failed to bring more protection to data citizens was the Rojas-Lozano vs. Google/reCAPTCHA case in the Massachusetts State Court. Gabriela Rojas-Lozano claimed that Google was profiting off of unpaid labor by asking for people to type two words that were displayed as distorted images (Harris 2015). She says that the first word proved she was not a robot, but the second word was unpaid labor. Google benefits from reCAPTCHA, as reCAPTCHA digitizes text images, creating profit down the line when they sell these products back to consumers. While Rojas-Lozano lost the case, her argument could've been improved. I'd argue that the key isn't in the second word in reCAPTCHA, but in the fact that data citizens don't know that when they do picture-to-words or word-to-picture matching tasks that they training an algorithm and also releasing cookies and data to Google without informed consent, a violation of GDPR. After discussing the technicalities in contract law with my friend Olivia Tyndall at Georgetown Law, one could argue that the contracts to use certain software-as-a-service are not valid because there is no informed consent for data citizens.

Another legal precedent that could be used to support an argument for the restructuring of the data economy is the 4th Amendment (4A) and the Third-Party Doctrine. The third-party doctrine permits the government to access information about individuals aside from 4A, which is supposed to protect a person from unreasonable search and seizure (Manning 2019). As technology has evolved, what information is part of our home and considered "private", and what information is allowed to be seen by the government is up for debate. Again, the key part of

informed consent or lack thereof is important to note here. People who keep an Amazon Alexa in their home may not understand that 1. They are being recorded by the device and 2. The government could potentially seize these recordings. They may believe that 4A fully protects them. How the third-party doctrine evolves legally over time will likely support a restructuring of the data economy, as without informed consent, many contracts will not hold when tested in court.

Lastly, there are already laws that exist to ensure that actors similar to technocrats in certain markets act as information fiduciaries, meaning that restructuring the data economy in internet policy would be following precedent in protecting data citizens (Balkin and Zittrain 2016). These laws exist for industries that are deemed necessary for how society functions, such as doctors, lawyers, utilities, etc. where doctors have to follow HIPAA laws and lawyers must follow attorney-client privilege. While they do not pay people for their data, they still must follow a set protocol in order to protect data, a right that is general knowledge. As Google and Facebook become indispensable parts of our lives, data citizens need special standards for how these corporations treat our data. The GDPR and the Data Governance Act are good foundations, but it is easy to see that internet companies are beginning to take up more crucial roles in our lives than most companies have before, and therefore need more nuanced structure to how they operate.

Economic

While there isn't a perfect economy that exists, the data economy, as is, has clear opportunities for improvement which would lead to more efficient outcomes for both data citizens and technocrats. This can most easily be illustrated by studying the case of the Google raters. There were over 10,000 raters who were contracted employees working up to 40 hours a week to ensure Google's algorithms were returning the intended results (Newitz 2017). These employees were subject to seeing horrific posts regularly without Google employee benefits because they were contractors. While this model is cost-effective for Google, this is clearly not the most optimal outcome for data citizens, raters, or even Google. Firstly, Google set out on an almost impossible task - there was no way that 10,000 raters would unanimously identify the best outcomes of all users. Secondly, these raters were exploited by their working conditions and

there was no accountability for Google. While content moderation is currently necessary to ensure safety on the internet, the people who actually have to moderate face horrific content everyday. Plus, paying people for their data allows Google to tie horrific information back to data citizens directly as they have a trail of the information's origination. Therefore, paying people for data reduces the need for content moderation as it creates accountability for data citizens for their own content, leading to more efficient outcomes overall (Iyer 2022). Thirdly, data citizens aren't getting an optimal product with the current structure that requires Google raters. Google was making judgements off of the information raters provided, but even they knew that rating was hard to get consistency with, citing that Google raters had to pass incredibly difficult exams in order to show they could deduce consistent ratings. If Google had instead paid data citizens some fee for their data and to verify the results they were getting, Google would have much greater accuracy of the data provided, and data citizens would receive not only direct compensation, but also a better product overall that was based on what they actually wanted, not what a rater and Google were optimizing for. Plus, paying data citizens for their data would generally generate more cash flows for ordinary people, bolstering the economy. This would not need to come at the expense of profits because 1. The product would be of higher value with better data, and 2. Technocrats already pay out such a small proportion of their value each year to workers that they could double the amount they'd pay and it'd make little difference to their profits. For example, Facebook pays out 1% of its value each year to workers vs. Walmart which pays out 40% (Posner and Weyl 2018). These technocrats clearly have the money to pay out for data citizens, but without any incentive they won't.

Political

    Besides the public's continued calls to regulate platforms, there are other political arguments to be made (in the US) to push for restructuring of the data economy. For starters, on the state-level we are seeing a continued decrease in available financial support for unemployed workers, increasing economic insecurity (Gwyn 2022). On the federal level, social security is running out and only expected to be payable in full until 2037, with the analysis still holding true today after the initial release from the Social Security Administration in 2010 (Goss 2010). The government doesn't have a way to resolve these issues, however the data economy could be a

saving grace. As I've stated, everyone has data, and data has value, therefore it should be allowed to be sold and generate income for the individuals that the data comes from. If platforms began paying people for their data, then people may not need to rely as much on unemployment or social security money, lessening the burden on the government. Plus, payments to data citizens for their data could in some way be taxed, increasing tax revenue for governments. If my argument from the economic section holds true as well, then technocrats would end up with higher-value products, resulting in higher profits and ideally, the government would get more in tax revenue from these companies as well.

In Practice:

While I've highlighted the reasons why the data economy needs to be restructured, there are still many ambiguities of how this would work in practice. Realistically, it is hard to determine the explicit value in dollars of any piece of data. The proposed models boil down to: a lump sum payment, a pay per data point, or a dividend model.

A lump sum payment could be paid to data citizens from a technocrat in the form of a direct payment, or through an auction-based model where companies bid for an individual's data or pooled individual data (Posner and Weyl 2018). An individual decides what data they want to provide to a technocrat, and that technocrat pays a flat fee for the data. In these scenarios, the data is valued by its potential value, and people can receive money right away. A direct payment gives immediate financial payment to the data citizen, but the individual may receive more money if they pool their data with others, resulting in a "sum greater than its parts" scenario. This is similar to how Nielsen ratings work. Nielsen pays a lump sum payment for a yearly contract with a household, and pays an annual bonus if the household continues its contract (Allec 2021). While this investment is easy logistically, it can end up decreasing the potential profits of a data citizen, as a company could potentially make much more money than what they originally predicted. So in this sense, a data citizen's value is actualized, but there's still a lot of room for shortchanging them in the lump sum scenario. Plus, once a technocrat has taken your data, it's possible that they still exploit the data and use it for more purposes than they originally said. Holding them accountable is still hard regardless. In the auction-based model of the lump

sum payment, there is some more legal sway here in a class-action lawsuit versus an individual versus the technocrat.

There is also the possibility of a pay-per-data point model, where an individual decides what data to provide to a technocrat and that technocrat pays for each data point based on how much value each data point is likely to bring. Here, valuations are again based on potential value, and it's possible that data citizens are shortchanged again, like in the lump sum scenario. Not only that, but bidding per data point is time-consuming and involves more charges and is therefore more costly logistically. However, it could prevent companies from asking for data that isn't necessary for the analysis, and only what is truly valuable to them. The infrastructure that's required here would make it easier also for data citizens to opt in and out of what data they want to provide. Overall this payment model is in some ways more developed than the lump sum model, but is clearly also more complex and not necessarily already implemented in practice.

Lastly, there is the data dividend model, where companies are required to pay a dividend out to all of its users that contribute data. Ideally, this dividend is updated regularly to reflect how valuable the data has been and how critical that data is to a company's value (Posner and Weyl 2018). This model seems to be the most optimal model, as it bases the value of the data on the actual value it provides, and also reinforces consistent communication and transparency between the data citizen and the technocrat. If data citizens want to opt out of providing data to companies, they can and will stop receiving regular payments as a result. This model also influences data citizens to continue to provide the most up-to-date information to technocrats, whereas the lump sum model provides no incentive for data to get updated. Plus, the infrastructure for this model already exists - publicly traded companies take value from people, in current practice money, and grant them a small share of ownership in the company, which also gives them some agency. It would be feasible to swap out exchanging money for consistently updated information.

Realistically, popularizing any of these modes of payment is incredibly important to bring more transparency and bargaining with technocrats to the mainstream. There are a few bargaining choices we have now, for example when looking at Google Drive vs. Microsoft Suite. Clearly, some people are willing to pay more to not give away their data and purchase Microsoft Suite, while others are willing to give away their data for free to Google Drive. Creating

structures in our system that force people to confront technocrats on what they're doing will hopefully popularize more rights for data citizens and also push people to make more informed decisions, again a more efficient outcome.

Conclusion:

Overall, while it will be difficult to implement, there are clear economic, political, and legal reasons we should transform the data economy so consumers can own their own data and be compensated explicitly. The law cannot strongly support the data economy existing as is, and the status quo is leading to inefficient business outcomes. There are also positive externalities politically that would arise from the creation and adoption of a more formalized data economy. While the way to go forward with this is up for debate, it's clear that the dividend model is likely the model that would work best in practice. In order to develop greater welfare for data citizens, technocrats, and the general economy, restructuring is clearly required and the current system will not stand.

Works Cited

Allec, Logan. 2021. "Nielsen Mobile Panel Review: What Is It and How Much Do You Make?"
     Money Done Right (blog). September 30, 2021.
     https://moneydoneright.com/fast-money/online-work/surveys/nielsen-mobile-panel/.

Arrieta-Ibarra, Imanol, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier, and E. Glen
     Weyl. 2018. "Should We Treat Data as Labor? Moving Beyond 'Free.'" *AEA Papers and
     Proceedings* 108 (May): 38–42. https://doi.org/10.1257/pandp.20181003.

Balkin, Jack, and Jonathan Zittrain. 2016. "A Grand Bargain to Make Tech Companies
     Trustworthy." *The Atlantic*, October 3, 2016.
     https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/.

Bureau of Labor Statistics. 2022. "Data Scientists: Occupational Outlook Handbook." September
     8, 2022. https://www.bls.gov/ooh/math/data-scientists.htm.

Burgess, Matt. 2020. "What Is GDPR? The Summary Guide to GDPR Compliance in the UK."
     Wired UK, March 24, 2020.
     https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2
     018.

Casilli, Antonio A. 2014. "No CAPTCHA: Yet Another Ruse Devised by Google to Make You
     Train Their Robots for Free." *Medium* (blog). December 6, 2014.
     https://medium.com/@AntonioCasilli/is-nocaptcha-a-ruse-devised-by-google-to-make-you-
     work-for-free-for-their-face-recognition-20a7e6a9f700.

Cohen, Julie E. 2019. *Between Truth and Power: The Legal Constructions of Informational
     Capitalism.* 1st ed. Oxford University Press.
     https://doi.org/10.1093/oso/9780190246693.001.0001.

"Datacoup." n.d. Accessed December 4, 2022. https://datacoup.com/.

Davenport, Thomas H., and D. J. Patil. 2022. "Is Data Scientist Still the Sexiest Job of the 21st
     Century?" *Harvard Business Review*, July 15, 2022.
     https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century.

Department of Economic and Social Affairs. 2019. "Frontier Technology Quarterly: Data
     Economy: Radical Transformation or Dystopia?" January 31, 2019.
     https://www.un.org/development/desa/dpad/publication/frontier-technology-quarterly-januar
     y-2019/.

European Commission. 2022. "European Data Governance Act | Shaping Europe's Digital
     Future." July 13, 2022. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act.

Goss, Stephen. 2010. "Research: The Future Financial Status of the Social Security Program."
Social Security Administration Research, Statistics, and Policy Analysis. August 2010.
https://www.ssa.gov/policy/docs/ssb/v70n3/v70n3p111.html.

Gwyn, Nick. 2022. "State Cuts Continue to Unravel Basic Support for Unemployed Workers."
Center on Budget and Policy Priorities. June 27, 2022.
https://www.cbpp.org/research/state-budget-and-tax/state-cuts-continue-to-unravel-basic-support-for-unemployed-workers.

Harris, David. 2015. "Google Duped Customers into Providing Free Labor, Suit Alleges." Boston
Business Journal. January 23, 2015.
https://www.bizjournals.com/boston/blog/techflash/2015/01/massachusetts-womans-lawsuit-accuses-google-of.html.

Iyer, Ravi. 2022. "Content Moderation Is a Dead End." *The Psycholgoy of Technology Newsletter*
(blog). October 7, 2022.
https://psychoftech.substack.com/p/content-moderation-is-a-dead-end.

Keith, James Felton. 2020. "Data Is Labor: Why We Need Data Unions." November 15, 2020.
https://www.coindesk.com/policy/2020/11/15/data-is-labor-why-we-need-data-unions/.

Kobie, Nicole. 2019. "Google's ReCAPTCHA Test Has Been Tricked by Artificial Intelligence."
Wired UK, March 25, 2019. https://www.wired.co.uk/article/google-captcha-recaptcha.

Manning, Grace. 2019. "Alexa Can You Keep a Secret? The Third-Party Doctrine in the Age of
the Smart Home." *Georgetown Law American Criminal Law Review*, Winter, 56 (0).
https://www.law.georgetown.edu/american-criminal-law-review/aclr-online/volume-56/alexa-can-you-keep-a-secret-the-third-party-doctrine-in-the-age-of-the-smart-home/.

Newitz, Annalee. 2017. "The Secret Lives of Google Raters." Ars Technica. April 27, 2017.
https://arstechnica.com/features/2017/04/the-secret-lives-of-google-raters/.

Pandya, Jayshree. n.d. "The Dawn Of A Data Labor Force." Forbes. Accessed October 26, 2022.
https://www.forbes.com/sites/cognitiveworld/2019/03/26/the-dawn-of-a-data-labor-force/.

Posner, Eric A., and E. Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and
Democracy for a Just Society*. Princeton University Press.
https://doi.org/10.2307/j.ctvc77c4f.

Powles, Julia. 2015. "We Are Citizens, Not Mere Physical Masses of Data for Harvesting." *The
Guardian*, March 11, 2015, sec. Technology.
https://www.theguardian.com/technology/2015/mar/11/we-are-citizens-not-mere-physical-masses-of-data-for-harvesting.

Scott, Laurence. 2013. "Who Owns the Future? By Jaron Lanier – Review." *The Guardian*,
     February 27, 2013, sec. Books.
     https://www.theguardian.com/books/2013/feb/27/who-owns-future-lanier-review.

Tartle. 2021. "Google Captcha Class Action Controversy - TARTLE." June 11, 2021.
     https://tartle.co/captcha-class-action-controversy/.

The Economist. n.d. "Data Workers of the World, Unite." Accessed October 26, 2022.
     https://www.economist.com/the-world-if/2018/07/07/data-workers-of-the-world-unite.

Thorburn, Luke, Jonathan Stray, and Priyanjana Bengani. 2022. "Is Optimizing for Engagement
     Changing Us?" October 10, 2022.
     https://medium.com/understanding-recommenders/is-optimizing-for-engagement-changing-us-9d0ddfb0c65e.

"TV Ratings." n.d. Accessed December 4, 2022.
     https://markets.nielsen.com/us/en/solutions/measurement/television.

"Wages For Facebook." n.d. Accessed October 26, 2022. http://wagesforfacebook.com/.

Wolford, Ben. 2018. "What Is GDPR, the EU's New Data Protection Law?" GDPR.Eu.
     November 7, 2018. https://gdpr.eu/what-is-gdpr/.

Yurkevich, Dennis. 2018. "Should We Treat Data as Labor?" *Medium* (blog). August 30, 2018.
     https://medium.com/@dennisyurkevich/should-we-treat-data-as-labor-4aaef549ab45.