Economic Impact and Feasibility of Data Dividends

Tarun Wadhwa

June 2020
EXECUTIVE SUMMARY

Personal data has fueled the rise of technology giants collectively worth trillions of dollars. As a result, there is a growing concern that users are getting a ‘bad deal’. In response to the imbalance in power and perceived loss of control, there has been an increasing interest in exploring the viability of instituting data dividends – a mechanism whereby companies share profits derived from the use of personal data directly with users.

Data is complex and its worth is contextual. The value of data is determined by how it is being used. At different stages in its production cycle, it has varying degrees of economic utility. Further, it’s not clear who “owns” data and what rights and responsibilities that comes with. It is often co-created by multiple parties as part of a digital interaction. Parsing who has claims to what is further complicated by advances in machine learning and predictive analytics. Data also comes in different types and formats, and therefore, from a technological and business perspective, much of it remains inaccessible and unusable.

There are many pressing arguments both for and against the concept of data dividends that are worthy of examination. If we were to be paid for our data, it could lead to fewer monopolies, clearer privacy rights, new sources of income, and unlocking large markets that are currently not possible with today’s structures. On the other hand, data dividends are quite difficult to implement, don’t solve the fundamental questions of consent and coercion at the root of the problems we have today, may slow down innovation, and could lead to its own inequality issues and externalities.

In theory, getting paid for “our data” and profiting from the information we create might shift the digital economy to a more equitable and productive orientation. However, in practice, there numerous obstacles and concerns over how this would work and whether it would deliver sufficient improvements over today’s architectures and arrangements. Without upgrades to the way we discover the value of data, new innovations in how users access and engage with their information, and major advancements in our legal understanding of what data is, the concept of data dividends is currently too impractical and unclear to derive major benefits for users.
INTRODUCTION

The ability of technology companies to collect, sort, and analyze our personal information has led to the creation of products, services, and platforms that have transformed the way we live. Their success has been so striking, in fact, that it is now raising questions about whether these companies are too powerful and if users are giving away too much value for what they are receiving. One solution to this situation that has been proposed is instituting a “data dividend,” a concept loosely borrowed from finance in which a company would compensate the user directly for the profits made from “their” data.2

In the United States, two recent developments from policy makers have renewed interest in the mechanism. In 2019, U.S. Senator Mark Warner (D-VA) and Josh Hawley (R-MO) introduced the Designing Accounting Safeguards to Help Broaden Oversight and Regulations on Data (DASHBOARD) Act, a bill to require the Securities and Exchange Commission to develop guidelines for how “commercial data operators” handle and price the information collected and derived from their users.3 Separately, Democrat California Governor Gavin Newsom called for his team to propose a “Data Dividend for Californians” in his 2019 State of the State Address, proclaiming that “consumers should also be able to share in the wealth that is created from their data” because “data has value and belongs to you.”4

Data dividends represent the next evolution in the debate over where and how we need to update our privacy laws and controls in response to advancements in technology. As Harvard University’s Berkman Klein Center Faculty Christine Borgman summarized, “paying people for their data is not new… it’s really part of a larger conversation about who pays for data, who controls it, who has access to it, and how you govern these relationships in a democratic society.”5

In this paper, we examine the feasibility of instituting data dividends, what benefits and challenges they will encounter, and whether it is likely to improve the balance of power between users and technology companies in the digital ecosystem.
TYPES OF DATA

According to Organization for Economic Cooperation and Development (OECD)⁶, “the environment in which traditional privacy principles are now implemented has undergone significant change” over the last several decades. There have been major improvements in the ability “to collect, store, process, aggregate, link, analyze, and transfer vast quantities of data.” That includes: the volume of data collected; the range of analytics providing insights into trends, understanding user movements, interests, and activities; the value of social and economic benefits that technologies enable; the changing nature of privacy; growing complexity of interactions; and global availability of personal data. As a result, the sheer size and breadth of what can be learned about individuals is far larger than at any point in the past, and it continues to grow rapidly.

These factors also complicate the considerations over what exactly “personal data” or “personally identifiable information” encompasses. While the OECD defines the former as “any information relating to an identified or identifiable individual,”⁷ in reality it can be quite complicated to determine what qualifies under this definition. In the past, personal data may have meant a driver’s license number or an address, but now online activity, user-generated content, or insights derived from algorithms can fall under this category. In addition, anonymity or obfuscation at the time of collection is no longer enough to ensure a level of privacy, as data can be linked back to individuals in unexpected ways, or combined with other data sets to draw revealing insights.⁸ The nature of discussion around data dividends extends beyond just personal data to “your data,” which in theory would include a right to ownership and control over all the ways an individual’s information can be processed. How far this extends remains undefined.

Data also has drastically different worth depending on what stage of production it is in. For example, consulting firm Accenture has a “Data Value Continuum” model that lays out five stages of data processing. It first begins with Raw Data collected from a source that has not (yet) been processed (e.g., sensor inputs). Then it becomes Processed Data, when it is stored, transferred, and put into summary form. From there it becomes Insights, when data science, mining, predictive analytics, and modeling are applied. That information is then taken to the Presentation stage, where it is put into intelligible formats and often used for decision making. The final form it takes is Transact, where it is in a form that can be made available to other companies and used for interaction with users. The value of data increases as it is processed through each stage.⁹ Another taxonomy proposed by the World Economic Forum divides data into three categories: individually provided (such as photos, social media posts, forms), observed (such as surveillance videos and location data), and inferred (such as credit scores or targeted advertisements.)¹⁰

Beyond classifications, it’s generally quite difficult to measure the value of an individual’s data. That’s because it is highly dependent on the context in which it is used. Data is also often combined with other information. As one analyst put it, “data is often valued within a relationship, but practically valueless outside of it.”¹¹ A list of a person’s restaurant reviews would be of little value to a company that manufactures thermostats, and vice versa for a person’s energy usage patterns. Moreover, it’s not clear that certain categories of data would exist without a company specifically prompting users to divulge it; search data might not be a category of much value if Google did not exist to cause us to undertake searches so often.¹² While it’s unclear who is getting a better deal in the relationship, it has to be noted that the technology company is undertaking the work to figure out how to derive value from these data streams. This gets even more convoluted when considering that personal data is often used at a large scale in machine learning models, where individual data input may be miniscule, but the insights derived from millions of samples of personal data may be quite valuable.
The data relationship is not restricted to just two parties. Not all data used to personalize or deliver services is even from the user itself—businesses can purchase data sets or share their raw data and analytics with other companies, purchase it from third-party supplies, record it from connected devices, or obtain it from personal-data intermediaries. As the OECD puts it, “subcontracting, outsourcing, evolving partnerships between organizations in value chains, behavioral advertising, and other emerging business models can add layers of complexity in determining responsibilities and identifying roles.”

Sharing of information is also hindered by the variance in types of data. There are multiple ways that data can be recorded, and that in turn limits how much utility can be derived from a particular subset of information. Several projects have sought to solve this issue, including the Ocean Protocol, which uses distributed ledgers, data exchanges, and smart contracts to enable the trade of personal information in a standardized format. Yet, as privacy researchers explained in a recent paper, there is currently “no universal method to track who shared what, with whom, when and for what purposes in a verifiable way to create an individual incentive for data owners.”

**PROS**

The point of data dividends is not to stem the use of personal information, but rather to drive it towards a more sustainable and equitable orientation. For businesses, personal data has led to better products and new markets; for users, it’s increased choice, convenience, and higher quality experiences. If data dividends were to be implemented, it’s possible that they could improve the status quo on information sharing in four key ways:

- **Increased user trust and participation**: One of the boldest arguments that proponents of data dividends make is this new economic arrangement will help align the incentives of everyone involved, which will ultimately lead to users being treated better. In their book *Radical Markets*, University of Chicago Law Professor Eric Posner and Microsoft Research Principal Researcher E. Glen Weyl argue that this will shift the current equilibrium drastically: “rather than being treated like passive consumers of the entertainments dished out to us by digital platforms, we would be honored as the suppliers of the data that make the digital economy work.” While it’s a difficult thing to measure, it could be the case that users are holding back from fully participating in the digital economy – evidence suggests there are growing worries about cyber security, data being weaponized against them later in life, and distrust between parties as a result of these ambiguities.

- **Compensation and income**: Over the long term, it’s likely that technologists will find new ways to gain value from sources of information by aggregating larger data sets and using improving analytics technologies to uncover deeper insights. It’s only right that users should financially benefit from this as well. The most valuable data sets may not be ones that are apparent to us today, and although evidence is limited right now, it’s conceivable that certain types of financial, health, and lifestyle behavior could be very valuable to specific companies or for specific applications. Posner and Weyl suggest that this could alter how we feel about potential job loss from artificial intelligence, they posit that “rather than the growing prowess of digital systems being seen as ‘Artificial Intelligence’ (AI) that would replace our jobs, it would be seen as a new source of well-paying jobs and income supplements.”
• Control and portability: If data dividends go beyond money and are paired with reforms such as mandating data portability, that could conceivably have a major impact on consumer lock-in and competition. Instead of being forced to use Uber for transit, a user could take their historical data, rewards points, and reputation score and apply it on Didi Chuxing, Lyft, or a new upstart that they wish to support. Yet we lack this level of sophistication in the understanding of our personal data because we don’t know the value of what we’re giving up yet. Once data can be accessed, moved, and controlled, it can then be donated. Users can share or contribute their data sets to the public commons, in order to advance research in ways that perhaps technically feasible but not culturally well-known yet.

• Enabling development of new markets: A story in Wired magazine explains the situation of Dr. Robert Chang, a Stanford ophthalmologist, who sought to build a machine learning model to scan eye image tracking to find conditions like glaucoma. Dr. Chang went to his patients, medical researchers, and donor registries, but was unable to build a large enough data set to perform his analysis. He’s now working to build a system where patients could be paid directly for their information — but in a world of data dividends, perhaps this would not be necessary. The scans could be stored in a machine-readable format, and the user could sell access to their data (either directly, or algorithmically) through an arrangement that the hospital has with different brokers. We currently do not know what applications and innovations are not possible because the required data to train models is not available. With the explosion of machine learning there will be new demands for huge amounts of information — but Chang’s story demonstrates instances where data accessibility hampers innovation. Information alone does not create new markets, but it can serve as the starting point. If rare data becomes abundant and accessible, it’s likely that researchers and entrepreneurs could find the applications and uses which would move humanity forward.

CONS

There’s little question that existing arrangements for information sharing are rife with problems. But moving towards a new model where users are paid for their personal data would almost certainly create its own set of issues. If data dividends were to be implemented, there is the chance that they could hinder progress on information sharing in four key ways:

• Difficulties of implementation: If a platform such as Airbnb uses a person’s prior interactions, preferences, and geographical information in order to improve their curation abilities, it’s unclear what exactly they owe the end user. Their aim is to return value for their data by delivering a better quality shopping experience, but that might be enough. If it’s determined that the user is owed money, what percentage of the company’s earnings are they entitled to, and how is that calculated? Data does not always fit neatly into classifications, and in reality, the questions of who owns and is owed what may be so cumbersome and heavy-handed that it will create more friction than users are willing to tolerate. Data will be increasingly generated by machines, many of which “have a direct or indirect human source,” which will make these questions even more difficult to work out.

• Consent and coercion: The current practice of “notice and consent” which dominates the Internet in the U.S. leaves technology companies in a position of power over users. By accepting their Terms of Service, users quickly sign away their rights — in a world where there are data dividends, this would not fundamentally change the way this works. Users can have greater legal protections, but if companies still set the rules of usage to their benefit,
then users can still be manipulated into giving data away. Legal scholar Elizabeth Renieris has argued that “that we live in a post-consent world and must find a way to adapt.” If companies now have to pay for this data, they could be entitled to go even further in their efforts to commercialize it. There is also the reality that anytime data is compensated, not everyone will be equally incentivized to participate. The poor are naturally in a greater state of economic vulnerability, so they may feel the need to increase the use of particular services or take actions that harm their choices in other ways, such as developing an addiction to exposing sensitive data.

- **Hampering innovation through complexity**: The added confusion over who owns what information and how they can use it will increase the cost of handling personal data. While this may cause companies to be more thoughtful in their approaches, it harms startups and new entrants who lack the teams of lawyers that their corporate competition employs. On the other hand, there is public value in the size of data empires that technology companies have built; while some might argue they are too powerful, there are virtually no other institutions in the world who can leverage their existing data sets to improve their products or enter new markets like today’s technology giants. Whether it’s Apple entering the banking industry or Google breaking into pharmaceuticals, regulating the use of personal data to a specific set of cases can close the door on developments that can conceivably improve our quality of life in ways that today’s measurements do not account for.

- **Targeting and externalities**: Not everyone’s data has the same value. In fact, one of the main reasons a company may want to solicit data from an individual is to balance out the diversity of their data set. With algorithmic bias a growing concern for business and government, there will be a push to make sure ‘training data’ is inclusive of all types of populations. That could lead to uncomfortable situations – where people’s age, race, class, and gender become the basis for why they are asked to sell their information, in an environment where legal protections are lacking. Then there are the people will attempt to game the rules. Just as we have seen ‘patent trolls’ try to abuse the legal protections around intellectual property, one lawyer suggests we may see the same thing eventually happen with data as well. Also, if data becomes something of value that can be sold directly, there will be incentives for people to cheat the system by submitting false, duplicated, or stolen data sets. This can not only lead to financial loss; it can affect the quality of decisions that are made using the information.
ANALYSIS

Based on the available legal regimes, technologies, and data architectures, the concept of data dividends doesn’t seem feasible. In its current conceptual form, data dividends will be costly to produce and are unlikely to shift the balance of power. However, that can change if three key obstacles are addressed: determining how to value data, building the ecosystem for personal data exchange, and settling the legal question of what data actually is. These are discussed in turn below.

Determining How to Value Data

For all the talk of the value of personal data, our measurements for what companies “owe us” are crude. They are frequently based upon the profits they make divided by the size of their userbase. Other times, they are taken from corporate filings about the amount of value derived per user. This reasoning misses a key point: it’s not cheap to do the sort of advanced analysis and modeling that makes personal data profitable. It often requires large scale technology implementations and advanced processing abilities that are prohibitive for non-experts to access. There are several mathematical approaches proposed for how to measure the value of an individual’s data, but they remain theoretical and untested in the real world. There are also emerging efforts to use blockchain-based tokenization to discover the price of goods, but this has not reached widespread adoption. For example, when exactly a user gets paid for their data can lead to different outcomes. If they are paid upon submission, their data can have one value, whereas if they are paid per use via an algorithm, it may be quite a different value. One can be paid up front, while the other can be paid over time. Data can gain value over time as it is used for new purposes, or lose value as it becomes stale and irrelevant. It may be the case for users that the value derived from companies reinvesting their profits back into improving performance outweighs the economic gain from a data dividend check.

Building the Ecosystem for Personal Data Exchange

It’s also not clear that users are ready to accept the responsibility and work that comes with managing their own data and digital relationships, in whatever form that might take. It’s one thing to receive payment from a platform you use, but what if you want to take that data and sell it to another company, as many have suggested as part of the vision of data dividends? One user reported that the amount of information Google and Facebook had on him individually was the equivalent of over 3 million Word documents. It’s not uncommon to hear stories of users requesting information from one of the major platforms and receiving hundreds of pages of printed documents in return. The average person would likely have little idea what to do with this, and may in fact resent the increased workload, calling into question the competitive benefits this policy might bring. Right now, companies cover the cost of securing that data and making sure it is free of sensitive information that could run afoul of the law. Users don’t have to think about who might owe them what and how they can collect it. In a sign that this issue only going to get harder to deal with, Intel estimates that a self-driving car will create 4 terabytes of data in an hour and a half of driving – figuring out what part of this is owned by the user, and what they should get paid for it, will be a monumental task. One day there may be unions, intermediaries, and virtual agents that help us manage this process, but until they are developed most of the grander benefits that might come from getting paid for your data will remain unfulfilled.
The Legal Issue of What Data Is

One of the largest barriers towards new models of data ownership, sharing, and profit-splitting that are associated with data dividends is the set of unsettled legal issues around whether data is an asset, capital, labor, property, or something entirely different. It’s also unsettled how selling data relates to privacy, which is a human right that cannot be commodified in the same way. In today’s system, if there is a dispute over what the worth of data is, there’s no authority which can settle this. It would conceivably start with some sort of ownership claims, which are a prerequisite to getting paid for what’s “yours.” As University of Oxford Visiting Fellow Jeffrey Ritter and University of Vienna researcher Anna Mayer argue, this can lead to much more, “once ownership is well-defined, then the attendant rights can be more precisely expressed — rights to access, license, transfer, modify, combine, edit, and delete data naturally flow from the control that ownership vests. In addition, both existing and new types of transactions can be more formally expressed (e.g., licenses, sales, transfers, processing services, storage services, analytics, and more).” Beyond just a check in the mail, there are all sorts of possibilities that open up, such as “sales transactions, licensing deals, joint ventures, downstream distributions and syndications of rights to access and use data, valuation for accounting and tax purposes.” Without this regulatory clarity in place, the idea of data dividends will be stuck in its infancy.

Overall, there is only a scarce amount of research available on how data dividends would actually work. The economic and social impacts they might have is largely unknown. As the concept grows in popularity and the problems of the digital ecosystem intensify, researchers, academics, and private industry should focus their attention on what users would want out of this exchange, what lessons can be drawn from the finance industry, how automation can be used to make value transfers seamless, how privacy-preserving machine learning technologies can improve the way transactions are done, whether the market pricing mechanisms in tokenization are applicable widely, and how these models can work across national boundaries.

CONCLUSION

The concept of data dividends signifies an advancement in the way we interact with and understand the value of personal data. As technological and analytical capabilities advance, we will be required to rethink our approaches and retool our defenses. In its current form, it’s not clear that data dividends would rectify the power imbalances that exist today between technology companies and users. While the value and applicability of data dividends may require further development before the concept is feasible to use, many of the tools, systems, and policies we develop along the way may be of great utility to advancing our digital ecosystems.
ENDNOTES

1 Tarun Wadhwa is Founder & CEO of Day One Insights, a global technology advisory firm, a Visiting Fellow at Emory University’s Department of Political Science, and a Visiting Instructor at Carnegie Mellon University’s College of Engineering. He is the author of the upcoming book, Identified: The Digital Transformation of Who We Are. He can be reached at mail@tarunw.com.


