



IARIW – CIGI 2023

IARIW – CIGI 2023

Thursday, November 2 – Friday, November 3

An Inquiry into the Production of Data and How it Creates Value Through the Ambient Economy

David Eliot
(University of Ottawa and The Pierre Elliott Trudeau Foundation (PhD
Scholar))

Paper prepared for the Conference on The Valuation of Data November 2 – November 3,
2023

Session 3: Methodologies of Data Valuation - II

Time: Thursday, November 2, 2023 [2:15-3:15PM EST]

An Inquiry into the Production of Data and How it Creates Value Through the Ambient Economy

David Eliot

University of Ottawa PhD Student, Center for Law Technology and Society
Pierre Elliott Trudeau Foundation (PhD Scholar)

ABSTRACT

This text asks foundational questions about data, and its origins of production, in order to construct a new model to understand operationally how data is created and commoditized to create value. Specifically, I position data as an inherent product of surveillance, resulting in a unique co-creation / co-ownership conundrum that must be resolved by legal mediating mechanisms for data to become an economic instrument in a market economy. Through my analysis I find that data is often not what is being traded as a commodity; instead, it is often the use rights to data that is commoditized and traded.

INTRODUCTION

Data has become an undeniably valuable resource, leading to claims such as it being the 'new oil' (The Economist, 2017). However, the properties of data as a resource remain unclear. This paper proposes that to accurately understand how data creates value, we must begin with a more thorough conception of data as a subject, which can then be extended to understand how data interacts with economic systems in order to create both 'use' and 'exchange' value. The goal of this paper is not to provide a methodology for assigning a specific value to data, but instead a theoretical framework for understanding how value is derived from data in a market economy. This theoretical framework may be utilized to produce future methodological tools and to analyze and contextualize current motivations and effects of data regulation such as the EU General Data Protection Regulation of Canada's Bill C-27.

In order to examine data as an economic subject, this paper begins with a theoretical exploration of how non-rival goods are given scarce form within economic systems designed to allocate scarce goods. I refer to the system through which non-rival goods are introduced into the conventional scarcity economy as the Ambient Economy. The text then pivots to an exploration of data as a subject, drawing on interdisciplinary literature to construct a holistic conception of data as a subject. Next, data is examined as an economic object within the market system, exploring the different forms it may take to create value. With an understanding of data as an economic object within the market economy established, I refer back to the previously explored ambient economic structure through which data is conceptualized as gaining its economic form, producing its value within the market system. In this section I establish, with support from other data scholars, that the valuable economic object observed is often not the data itself, but use

rights regarding the data. With these pillars established, the central question of this text, how data creates value, is directly addressed in section VI. The remainder of the text is dedicated to briefly exploring the practical implications of the theoretical analysis, examining how revelations regarding how data, and its derivatives, gain 'value' can be used to re-contextualize current policy debates and legislation regarding data privacy as inherently economic actions.

I

The Ambient Economy is a conceptual framework for the interconnected network of systems, agents, and institutions that provide non-rival goods a scarce form and enforce their scarcity. In short, ambient economic systems allocate goods that are non-scarce. Ambient economics is a conceptual way to visualize how non-rival goods are transferred into conventional scarcity economies so they may act as economic objects. The non-rival goods which are subjects of the ambient economic systems may be referred to as 'ambient goods'.

Conventional economics are primarily, and historically, engaged with scarcity-based economies and systems. The term 'economy' is often defined as a "mechanism for allocating scarce resources" (Mankiw and Taylor, 2014). Any conventional economic system, at its most basic level, is a system to address the three central questions: what will be produced, how it will be produced, and how will the output be distributed. Economic systems are necessary due to the fact that goods are finite, desires are infinite, and therefore there needs to be some order to how they are distributed. The capitalist system –in its purest form– relies on market forces to determine what should be produced based on the market's demand. The goods are then distributed based on who can afford to purchase them. In a communist economic system, it is the central government that decides what goods are produced with the resources available. Further, the central powers will decide exactly how the scarce goods produced should be distributed to the populace. But what about non-rival goods, goods that are practically infinitely reproducible; whose reproduction does not diminish others' ability to use such goods?

A pragmatic and simplistic approach would be to assume that these goods are not governed by the conventional economy, as they are inherently non-scarce. But this is obviously not the case as non-rival goods are constantly traded within, and governed by, conventional economic systems. One prudent example is the commodification of stories. A story is a non-rival ambient good as it can be retold infinite times and through infinite mediums. Further, one subject's act of telling the story does not diminish another actor's ability to tell the same story. However, in the United States¹ legal mechanisms have been put in place to make some stories scarce. You cannot make and sell your own copies of the Hunger Games books. This is because the right to produce and profit from the story of Hunger Games is protected by copyright law, making it scarce. Telling

¹ I am utilizing the United States as it provides a clear example. The example can be applied to countless other IP protection systems.

the same story does not diminish the story itself, but does diminish the original author's social right to profit from the story. You further cannot write your own slightly different version of the story without potentially infringing on the right to story derivatives awarded to authors by US copyright law (Boyle 2008). I construct the process of copyright as being a mechanism of the ambient economy of the United States.

Ambient economic mechanisms are tasked with producing artificial scarcity for non-rival goods. The scarcity examined takes a socio-legal form, as the resource or good itself is not scarce, however, the creation of a social contract through the legal system makes it scarce through collective adherence to the legal system. This can create geopolitical tensions when the legal forms accepted in one regulatory/legal region are not respected within another (Mandel, 2006). In the case of patents –which create scarcity for invention ideas, another non-rival good– there are international agreements and governing bodies in place to govern, regulate, and ideally enforce the concept that a patent reserves the right to an invention for the patent holder, creating scarcity for that idea (Boyle 2008). Here we can observe how ambient economic systems, similarly to conventional economic systems, can stretch across geographic and political space, connecting similar economic environments.

The form of a given ambient economy is inspired by its response to constraints. Whereas the conventional economy is required and designed to respond to internal constraints, particularly the constraint of scarcity of resources, the ambient economy takes shape and allocates in response to exterior social constraints. Many of these forces are produced by the ideological form of the economic system that similarly shapes the conventional economy. For example, the aforementioned example of the US copyright exists as a result of the same ideological forces that shaped capitalism itself; a desire to motivate innovation and competition via individual reward (Boyle, 2008). In a hypothetical hunter-gatherer society, there was likely no need for an official copyright system, as the ideological motivation of the economic system may have been more communal. Additionally, there are numerous forces that may act upon an ambient good to make it scarce, other than economic motivation. These forces may include social, or human rights claims, such as the right to privacy. Such rights claims become extremely important restraints when discussing the process of synthesizing data as a scarce resource, and allocating rights to it. If there are no forces acting upon the non-rival good, it may enter the conventional economy as a 'free good' which can be freely consumed by any subject with little to no acquisition cost².

Although it may be tempting to treat the ambient economy as part of the conventional economy, I hold that it is worthy of its own consideration and examination as a subject of inquiry. It is possible that the collection of structures I call the ambient economy may have gone generally under-recognized as a unified system with its own purpose and logic, as its integration into the

²After the copyright period has elapsed, a story may enter the public domain, losing its scarce form and becoming a free good.

conventional economy historically has been relatively straightforward. As such, constructing it as its own system would be needlessly tedious. If a non-rival good requires a scarce form, one could easily be provided via a legal system such as a copyright or patent, which is then allocated to a specific agent –such as a patent inventor, or the artist that created an artwork– as a good with scarce form that could become an economic agent within the conventional economy. However, the growing prevalence of data as an ambient resource has created novel issues within the process of scarcity formation and allocation. It is data's behaviour in the ambient system which prompts the necessity of our attention. The difficulties of assigning data scarce form are amplified by the fact that data has become an exponentially important resource for the functioning of our society at large and a primary driver of profit for businesses. As will later be explored in this text, it is data's scarce form, provided by ambient economic institutions and mechanisms, that construct its exchange and market value.

II

Data

In order to examine data as an economic subject, it is vital that we first construct a robust understanding and definition of the subject that is being discussed. An exploration, and definition of data must, in my opinion, cover two concepts. First, data as an object must be defined. Second, an understanding of the nature of data's existence and where it comes from must be developed.

Numerous texts dealing with data as a central subject of their analysis, fail to offer any substantial definition for data (i.e. Batra 2008; Buckland 2017; Andrews 2012; Sumpter 2018; Mansell 2009). In many cases where data is defined, the definition is brief and under-explored. For example, Couldry and Mejias simply define data as digital information produced about humans via a mediated interaction with technology (Couldry and Mejias, 2019). Couldry and Mejias' definition highlights another common issue concerning current data definitions: the tendency to center humans. Specifically, Dutch-Brown et al. (2017) and Couldry and Mejias' (2019) definition define data as something about humans, excluding all non-human data such as animal data, environmental data, or machine data. This issue can be solved by dividing data into two categories: machine data, and human data, as done by Dutch-Brown et al. (2017). This distinction is understandable for legislative reasons, however, I hold that it is problematic as a starting point, as it positions machine and human data as being fundamentally different objects. It is my position that we must first define data in its broadest senses, and produce an understanding of its properties, before narrowing our scope to examine data about specific socially defined subjects.

For my conceptualization, I leverage the definition of data coined by Tom Stonier (2012) to form a more holistic conception of data. Stonier defines data as a collection of perceived facts. Within data, each individual piece of data is colloquially referred to as a datum. Within this definition,

each datum represents a perceived fact. A poetic articulation of this concept emerges from Tang (2021) who refers to data as a ‘frozen memory’. Stonier's (2012) definition has further support from intellectuals, as it mirrors what appears to be the dominant understanding of data within the information science community (Zins, 2007). Further, the concept of data as perceived facts align with the operational definitions provided in several computer, and data science textbooks that I examined. Zhou (2021) defines data as a record, whereas Kelleher and Tierney (2018) position a datum as an abstraction of a real-world entity.

From this definition, we can narrow our scope to examine more precise types of data. Human data is a collection of perceived facts about humans. Personal data— a collection of perceived facts that includes information regarded (via a socially constructed definition) as personal. Digital data is a collection of perceived facts, stored on a digital medium. Archival data is a collection of perceived facts, stored in an archive. Library index cards are a data set about the library, a collection of perceived facts regarding the location of the books on the shelves. Each individual card, a datum—an individual perceived fact— within the set. Note that the fact is perceived, as a patron may have inadvertently placed a book back in the wrong location. This does not change the datum's existence as a datum, as it is still perceived to be true. Biased, or inaccurate data, is still data.

The next step in understanding data is to assess where it comes from. The concept of data's production, much like its definition, is lightly explored. Surveillance scholars such as Zuboff (2015), Fuchs (2012) and Sadowski (2019) discuss data as being produced by the subject that it is extracted from, however, they do little to examine the mode by which the data is produced by the subject. Batra (2008) additionally positions data as something that is collected from users, although again does not elaborate on how users produce it. Couldry and Mejias note that it is produced via a mediation, but do not elaborate on what this mediation entails (Couldry and Mejias, 2019).

Some clarity in this discussion is provided by Lev Manovich (Lev Manovich, 2001 as cited in Gittelman and Jackson, 2013) who conceptualized data as something that is generated. Gittelman and Jackson (2013) expand on Manovich, stating that an event must be imagined as data before it can become data. I expand on this concept, positioning what Gittelman and Jackson (2013) refer to as ‘imagination’ and Manovich (2021) as generation, to be a synthesis produced by surveillance. *Specifically, I proposed that data is inherently the product of surveillance.* Marx (2002) presents a similar thesis, interlocking data production and surveillance, by defining surveillance as “the use of technical means to extract or create personal data” (Allmer, 2012). This proposition should not be confused with Schneier's (2013) proposition that data is collected through surveillance, as their analysis relies on the data existing prior to the act of surveillance.

I hold that all action has the potential to be data (Eliot, 2022). We (as well as every other material object) are constantly in a state of creating data potential. This is not data itself, but the potential for data to be created— one half of the equation. For data to be created, an action must first be imagined as data before it can become data (Gittelman and Jackson, 2013). Once it is imagined

as data, an actor may surveil the potential data, recording it to a medium, synthesizing the potential data, into data. This may be conducted via a complex technical process such as web cookies, or can be as simple as a kid recording the strike counts for a baseball game on a scorecard. I invoke surveillance, instead of observation, as the act is not passive. The surveillor must actively record the perceived fact.

Note the unique feature of data that this conception reveals. Data is not natural-- but produced. However, it is not produced by a solo actor. There are two agents involved in the synthesis of data-- the surveillor, and the surveilled subject (Eliot, 2022). On occasion, these two subjects may be the same agent (i.e. if one keeps track of the number of tacos one eats in a year, they are surveilling themselves to create data). However, more often than not, the two agents are not one and the same. For example: myself and Facebook, or a self-driving car and the other cars on the road.

III

The question that this text is grappling with is 'how data creates value'. With a definition of data established, and a theoretical framework for how data is produced, we have inched closer to an answer. However, it must also be established what kind of economic theory of/about data I am trying to produce. Specifically, I seek to avoid constructing a narrow use-specific theory.

Use-specific theories can be represented by the likes of Zuboffs (2019) *Surveillance Capitalism* Fuchs (2011; 2019) *Web 2.0 and Data Capitalism*, Srnicek (2016) *Platform Capitalism*, Andrejevic (2010; 2019) and Couldry and Mejias (2019) *data colonialism*. These texts ground themselves in a specific application of data. For example, Zuboff (2019) is interested in how data is used to produce value/wealth through its utilization in advertising, and the production of 'behavioral futures'. Texts in this category commonly examine how economic systems and actors exploit data's value creating potential for economic gain. Although extremely important--this should not be taken as diminishing the value of these texts-- I note that such texts place data within the context of the capitalist economy, seldom if ever exploring data outside of the capitalist economic system or its inherent traits that are perceived to be exploited. Because of this, such texts are not truly exploring the forms by which data creates value, but instead the relationship between data and the economic system and agents of capitalism. As Zuboff herself notes, the subject of her work is not the data itself, but the behavioral futures, which are a derivative commodity produced from the data (Zuboff, 2019).

The goal of this text is to explore data as a more abstract subject that exists both outside, and within the capitalist system. My focus is on how data can be understood in general terms that can be applied to a plurality of current, past, or future economic systems. An example of a similar analysis can be found in Sadowski (2019) *When data is capital: Datafication, accumulation, and extraction*. Sadowski's text positions data as the key economic driver of the 21st century, and desires to uncover the forms it takes within the market economy (i.e. is data capital or a

commodity). Although the analysis is grounded in the market system, it is uninterested in the specific ways that data produces value within the market system (such as marketing, AI services, managing etc), and instead focuses on how data acts as an economic good.

A clear way to conceptualize this separation is that a specific analysis looks at how data creates value, and a general one looks at how data itself gains value. It is important to note that by value, I am referring to economic value. Again using Zuboff (2019) as a foil, in her work it is the behavioral futures that produce economic value for Google and Facebook. In Srnicek's (2016) platform capitalism, it is the platform itself that generates the profit. When examining the use of data in financial markets, it is the selling of, or leveraging, information generated from the data that produces monetary value. Data is being used to create value, but it is not the value of the data as a subject or resource that is itself being primarily analyzed. Alternatively, Sadowski (2019) is interested in how data as a subject acts as an instrument of value, specifically whether data acts as capital or as a commodity. Sadowski (2016) utilizes Bourdieu's conception of economic capital to argue that data exists as an economic entity that exists, and has value, exterior to the knowledge about the world produced from it. Data then itself acts as a form of capital, similar to a factory machine –although Sadowski identifies numerous differences– that can be leveraged to produce valuable goods. An important segment of Bourdieu's theory of economic capital is that economic capital is institutionalized through property rights (Sadowski, 2016).

For this exploration, I will also be grounding the understanding of data's value-producing capabilities within the market system, although the general concepts can be applied outside of it. This is for two primary reasons: first, it provides an easy reference for the abstract concepts that will be developed, and second, it provides more direct utility for the concepts developed as the market-based economic system is currently the globally dominant economic system.

IV

Building from Sadowski, it is important, in order to understand this paper's central question, to develop an understanding of the economic form data takes. I establish the economic form for the purpose of this text through Callon's (1998) definition of an economic instrument: a subject that can be calculated into economic formulas. Sadowski (2016) positions data as a unique form of capital, and not as a commodity due to its use as a subject that is used to produce commodities. He further notes that it takes on some of the features of financial capital, as the new commodities produced from data typically in turn produce more data to be used to create more commodities. However, depending on the operational definition of 'commodity' it can also be possible to position data as a commodity itself.

Defining the term commodity can be difficult, as it may mean different things, and be used in different ways by diverging schools of economic thought. For example, one may quickly shoot

down the idea of data being a commodity, as it is non-fungible. In many neoclassical definitions of a commodity, the commodity is a fungible good. This can be observed in the World Trade Organization rules for commodity markets, which require commodities to be 'alike' (WTO, n.d). However, this rigid view is not universally applied. In his Nobel prize winning economic research on social choice theory, Kenneth Arrow constructs information as a commodity (Arrow, 1973; 1978). This is despite the fact that information lacks several of the defining features typically associated with commodities such as fungibility (Arrow, 1973; 1978). Packets of information, or information itself, are inherently 'not alike'. Instead, Arrow conceived of a commodity as a subject with value that can be traded amongst individuals (Arrow, 1973; 1978). This parallels Karl Marx's (2019) definition of a commodity as an "external object, a thing which through its qualities satisfies human needs of whatever kind". For Marx (2019) the value of a commodity is produced by the labor used to create it. This differs substantially from classical economic theory in which value is produced from scarcity and the forming of a market price equilibrium. Notably in Marx's conception, there is space for commodities not formed through labor such as natural resources, however, they must be assigned property rights, establishing clear ownership, in order to take the form of commodities. Such goods have been referred to by Karl Polanyi as fictitious goods, as they are created outside of the sphere of production, and assigned legal form in order to become economic subjects within the market (Jessop, 2007). Callon (1998) further develops the theoretical conception of commodity, stating it is a good which may be traded on a market, subject to its calculatable equations. He built on this conception, examining how a good becomes a commodity in market systems, concluding that it must be disentwined, and alienated from its previous attachments so that it may produce value and be traded as a market commodity (Callon, 1998).

As noted by Sadowski (2016) Marx does not consider an object/subject itself to be capital, as capital is not a physical thing, but a relationship between the capitalist and commodities. In Sadowski's reading of data, it becomes capital because it is used to create surplus value that can be leveraged for more profit. However, it is my position that data can act as a commodity, as well as being capital, within both a classical economic, and Marxist framing (although it will never be fungible).

It is important to state upfront that Sadowski's (2016) proposition of data as a form of capital is accepted within this analysis. Although Sadowski positions his analysis in Marxist economics, I hold that the analysis is also applicable to classical economics.

Further, I suggest that data separates itself from IP such as a copyright, which is typically considered to be an intangible asset, due to the necessity of retaining data in a material form in order to utilize it to produce commodities. Where commodities are typically produced in reference to intangible assets such as IP –the IP for Toy Story is not an active ingredient in the production of a Toy Story film, but instead is the legal right to do so– the physical presence of

data is necessary for the production of subsequent commodities. Data can additionally be conceptualized as a commodity wherein the data itself is exchanged as an economic good. For example, I may purposely and independently produce voice recordings that I sell to a data broker, who may then package my recordings with others into a marketable data set. Similarly, a hobby meteorological researcher may buy a data set regarding global tidal patterns in order to run a simulation for a research project. In such a case the researcher is consuming the data in order to achieve an objective which is not necessarily to create a new commodity. In such cases the data, or datasets being exchanged are being exchanged as commodities, however, the same transactions would take on a different form if being sold to an entity which repetitively utilized the data to create products such as behavioral futures.

V

Data and the Ambient Economy

Data behaves differently than many goods within the ambient economy due to the nature of its production. Data is non-rival as it is infinitely reproducible, and one actor's use of data does not diminish another actor's ability to use it. An example of this is the Israeli government providing national health data to Pfizer in exchange for expedited and privileged vaccine access (Zion, 2021). Besides potential ethical or privacy concerns, the Israeli government was able to give Pfizer access to the data for their use, without diminishing the government's ability to use the same data. However, data is uniquely different from other non-scarce or non-rival goods, as it is not as much produced as it is synthesized through the collision of two actors, who both have recognizable rights claims to the data.

As discussed in Section II data is produced through the action of surveillance. It inherently requires two subjects, a surveilled actor, and a surveillant actor. The surveilled actor produces the potential for data, and the surveillant actor records it as a collection of perceived facts. If the surveilled actor is inanimate and or owned by the surveilled actor, assigning rights could potentially be relatively straightforward. However, once the surveilled actor becomes a subject with the ability to make rights claims over the data, a new form of mediation is necessary.

Relying on Callon's (1998) explorations of how objects become economic instruments in a market economy, we find a unique issue with data. Specifically, since data is formed via a synthesis between two agents, it must be disentwined from the surveilled agent in order for it to be introduced into the market as a commodity by the surveilling agent (Callon, 1998). The surveilling/recording actor must establish clear legal ownership over the data. The process of disentanglement does not necessarily mean a severing of all links, as contractual stipulations such as how one may use land that is transferred as a commodity, do not impede the transfer right to the land.

In a hypothetical State where there are no data protection or privacy laws, the surveilling actor may simply be awarded the right to claim ownership over all collected data as if it is oil extracted from the earth. The data, packaged, and assigned a market value could be freely sold on data markets. This example, however, raises another important question: what is being sold? The data or something else.

Here we must draw an important distinction. When dealing with data as a commodity, completely disentangled from the subject it is synthesized with, there are two primary market forms it may take. First is the data itself as an economic subject, and the second is as a usufruct/use right. When data is traded on the market as its own commodity, ownership rights must be fully transferred. The surveilling actor must be disentwined from their ownership stake in the data, seeding control, such as the rights to sell the data over to the buyer³.

The second form data can take in such a sale is as usufruct. Usufruct, also known as use rights, refers to the idea in property law that a subject may be granted a limited right to use another subject's property under set conditions (Alston and Mueller 2008). In such a situation, the seller isn't giving up their ownership rights to the data, but selling the right for another party to use the data. The buying party may have restrictions placed on how they use the data, or may be able to freely use it for whatever purposes they desire; however, *they lack the ability to claim legal ownership over the data*. As the purchasing actor may have a hard drive containing the data, and use it numerous times to create countless products and other commodities, the use rights themselves take the form of capital. Notably, it is the use rights that take the form of capital, and that were the commodity transferred, not the data itself.

circling back to the hypothetical scenario, we do not in reality inhabit a society with no established privacy or data rights. Although not universally enforced, the right to privacy is enshrined in Article 12 of the *United Nations Declaration of Human Rights* (Universal Declaration of Human Rights, 1948). Western legal scholarship has further established the difficulty, if not impossibility of fully seeding ownership over personal data due to the existence of privacy as an inalienable right (Litman, 2000; Samuelson, 2000) Further, most Nations or regulatory regions have their own privacy and or data laws. These laws and rights govern the relationship between the surveilling actor, and the surveilled subject that synthesized the data, and construct the form through which data may become an economic subject or commodity. In short, data is being constrained by the exterior force of rights.

³ Since data is non-rival, I conceded that the agreement could theoretically involve the original owner (the surveilling actor) maintaining ownership rights of an identical packet of data. However, the transaction would extensively create two distinct economic entities, and the surveiller would be disentwined from the ability to control the data packet sold to the buyer. Notably, this does not include legal agreements that are made in the transaction, such as non-competition clauses, or agreements that the data can't be used for specific purposes.

Not only is data as an economic subject being constrained⁴ by these rights, but the form these rights take, or are interpreted through, is constructed through the ideological position of the regulatory regime. For example, data rights take different forms in the United States, from the European Union. In the United States, where ideology is more business-friendly, the surveilling actor is granted more lenient rights over the data (Eliot and Murakami Wood, 2022). However, in the EU, the GDPR presents stricter regulations that grant the consumer greater rights over their data, its use, processing, and sale, as well as the right to refuse to be surveilled (GDPR, 2019). At this point, it would become far too easy to get bogged down in the specifics of how different regulatory regimes mediate this relationship. It is important to recognize instead that these regulatory relationships restrict either party from claiming ownership, and structure the economic form personal 'data' will take.

I put data in quotations in the concluding sentence of my previous paragraph, as it is my position that in these cases where the data is incapable of being legally disentwined from the surveilled subject, it is not the data at all that is entering the conventional economy as a scarce good— but the usufruct (use right). This is heightened by the fact that it is generally accepted that ownership of one's personal data cannot be simply signed over via a 'terms of service agreement' due to the inalienable rights associated with privacy (Litman, 2000; Samuelson, 2000). This is in part because it is generally accepted that ownership rights cannot legally be established for personal data (Scassa, 2019). What is signed over is instead an agreement to allow the surveillant party to use the collected data in specific ways, such as for advertising, creating a contractual usufruct regarding the synthesized data. Scassa has noted a similar phenomenon stating:

although the personal data economy is burgeoning, it appears to be based more on contractual models than on any underlying ownership right in personal information (Scassa 2019)

The surveilling party may also retain the right to provide the data to other entities, however, they do not have a property ownership of the data. In the case of Google, their terms of service directly state that they do not have the right to sell the user's data (Google Safety, n.d). However, even without this right, Google is able to extract immense value and profit from the data via their usufruct.

The usufruct may act as a commodity, as the surveillant party may, if their rights allow it, sell other actors the ability to access and consume the data. Further, the surveilling actor is selling the commodity of access, making the usufruct act as capital. If the surveillant actor is not selling access but is using the access to produce other commodities, such as AI systems that are then sold as data, the usufruct is again acting as capital, with the sold AI systems being the produced commodity

My arguments thus far can be operationally stated as such:

In the conventional economy...

⁴ This should not be read as a critique of these constraints. I fully support the right to privacy.

1. A good must have clear legal ownership to be traded as a commodity on a market
2. The value of a commodity is created by scarcity

In the ambient economy....

1. Data is infinitely replicable
2. Data does not freely exist as a resource to be extracted.
3. Data is the product of the surveillance of information (which is also infinitely replicable). It is created when an action is recorded as a statement of fact
4. There are two parties involved in the creation of data, therefore both parties have a claim to the data.
5. It is the job of the ambient economic system to reach a compromise of allocation between these
6. two actors so that the resource can be introduced into the scarcity economy (market economy).
7. Full ownership can be given to either party, however doing so negates the other's claim to the production of the data, but does not completely nullify the claim itself. The claimant can still argue that they have a claim to the data, using other social contracts that are integrated into the economy such as that of human rights.
8. If ownership is not assigned, rules must be determined that govern the relationship between the two parties.
9. Data can only be capital or a commodity when full ownership is granted; otherwise it is access to the data that is capital or the commodity. Access has stipulations created by the market's rules of compromise.
10. Access to data, or the ownership, can be used to produce commodities.
11. Therefore, Data can act as a commodity or capital, if the scarcity economy (the market) rules allow for a given datum to be exclusively owned, and the datum is exclusively owned. Otherwise, it is the Access to the data that is the commodity or capital. This is most prevalent with personal data, where claims to data rights are most strong.

(Eliot, 2022)

VI

How data creates value

Let's return now to the central question of this text; how does data create value? So far we have traced the economic life cycle of data, from being imagined, to surveilled, to synthesized, mediated into a legal and scarce form, and commoditized. A vital revelation –although this text is not the first to note it– is the fact that the economic entity typically referred to as data, that is leveraged to create value, or traded as its own instrument, is not data at all; but instead rights associated with data. What I have referred to as 'use rights' or usufruct. With the method of data commodification established, it is now possible to tackle the question of value.

In traditional Marxist theory, which is often used to conceptualize data in the social sciences by thinkers such as Fuchs (2010; 2012; 2019), the value of data is derived from the labor contributed to its creation. The surveilled subject is conceptualized as engaging in labor which is exploited by the surveilling actor when the data is extracted (Fuchs, 2010). In order to make this conception work, the idea of the 'prosumer' is invoked, positioning the surveilled subject as doing labor through the act of consuming (Fuchs, 2010; 2012). The positioning of data's value extending from labor is utilized to form arguments regarding the exploitation of the surveilled subject during the data extraction process.

I reject the premise that the value of data is created through labor on several fronts. First, conceptualizing the user's actions in the data synthesis process as labor directly contradicts Marx's own definition of labor. Marx defined labor as such:

Labour is, in the first place, a process in which both man and Nature participate, and in which man of his own accord starts, regulates, and controls the material reactions between himself and Nature (Marx, 2019).

For Marx, labor is something that is intentional. The production of data potential that is synthesized into data is certainly not always of one's own accord. Further, such a construction leaves no room to understand how data about non-human subjects gains value, for example, data on average global temperatures. Secondly, in an effort to conform to traditional Marxist theory, the suggestion that the action of creating data potential is labor devalues the concept of labor in its entirety. Let us ask, should all action taken by a human be deemed labor before it is surveilled in order to produce data? Or does an action only become labor once it is surveilled? If the former, are we then in a constant state of suspended labor? If the latter, are we not then in a state of constant labor due to the existence of global imaging satellites? The thought of being in constant labor is not only exhausting, but unnecessary, especially as Marx leaves room in his writing for other forms of exploitation of the proletariat other than labor.

It should be noted that under Marx's definition, it could be conceived that it is actually the surveilling actor that is creating the value, as they are the ones doing the intentional labor of creating the surveillance apparatus. This is reflected in some current experimental conventions for the valuation of data. Specifically, Statistics Canada values their data based on the labor costs to procure/produce the data (Statistics Canada 2019).

It is my position that the use value of data is derived from its potential to be combined with other data to be processed, in order to satisfy a human desire. The use value of an individual datum relies on the trustworthiness of the surveillant actor. Trust that they are truthfully recording an abstraction of the world, or of the subject the data is said to be in reference to, and not a fabrication. The value comes from the creation of a 'true' perceived fact. At the universal level, stripped of all social meaning, all perceived facts are equal in use value. Each individual datum contains the same potential to be aggregated to create something more to fulfill a human desire.

The surveilled subjects, whether they be a rock, an ant, or a billionaire, are all equal in value. The value they add comes simply from existing in a state to be surveilled. But we do not live in a universe stripped of social meaning. Therefore the use value must also be understood as being produced through the potential use to fulfill human desires that the abstraction of the surveilled subject, as represented through data, has. Therefore the practical use value of data is in constant flux, depending on the usefulness of different data types for different data-consuming processes.

An economic value may be assigned to the use value of data via an attempt to ascertain the financial gain that could be derived through the processing of the data. This calculation will always be crude, as it is impossible to know what technological innovations will occur that increase the usefulness of data about specific subjects, or how the data may behave when combined with other data. Again I note that the use value is typically realized from the combination of data, which creates this unique value conundrum. Additionally, the use value of data is typically not derived from the data itself, but instead from the data's integration with traditional business practices (CPA Canada, 2021).

The economic value of data in a market system is derived from the market's belief of the potential use gain from processing the data. Since data is non-scarce, this equation becomes quite fascinating. If a dataset is made completely public, the dataset itself has theoretically no market exchange value. Trying to sell the same dataset would be pointless, as the buyer would receive no potential gain from the purchase. The economic valuation of the data is derived from the exclusivity of access to it— the use right— and the ability to leverage its use value. The economic value of an openly available dataset could then theoretically be formulated by establishing what the market price would be if the data was exclusive. Here we can observe the divergence of data's exchange value, and its use value, as if the data were to be public, it still has immense use value to those who use it to fulfill their desires, but it has lost its exchange value due to its loss of a scarce form.

When attempting to provide a value for data, we must be clear if we are evaluating its exchange value, use value, or a combination of the two.

With regard to use rights as an economic object, the results are parallel with slight divergence. A use right has a use value that can be determined by the potential the data has to create future profits. This value may be directly affected by the exclusivity of the use rights. If two actors have use rights to the same data set, they may create products that are in competition with one another, reducing the potential value of the data to either company. Notably in this scenario it is the use right losing value, as its potential to fulfill the human desire (make more profit) is reduced; **it is not the data losing value**. Thus, the market exchange value can be conceptualized as being derived from the expected economic gain purchasing actors will receive from their use rights. Again this is diminished if the use rights are made non-scarce. Pfizer would not have been

willing to give Israel privileged vaccine access if Israel was also going to make the use rights to the data public. Additionally, it is theoretically possible –although unconfirmable– that if Israel was offering this use right to multiple vaccine manufacturers, that the value return to Israel per transaction would have been lower, as part of the use value of the transferred data was the competitive advantage the ability to process it offered to Pfizer.

VII

The big question that must be asked is, why does this all matter? Up until now, this text has presented ample theoretical abstraction, to make seemingly simplistic points:

1. The value of data is determined by its ability to be combined with other data to be processed in order to fulfill human desire
2. It is often use rights, or access to data that is traded as an economic subject, and not data itself
3. The market value of data or use rights to it are derived from the perceived use value and the scarcity of the data. This could be described in basic economic terms as supply and demand.
4. It is the scarcity of the use right, or access to the data, that creates commodity value.

Although these points may appear simple, it is how we arrived at them that creates interest. Specifically, the discussion of how data, or use rights, are synthesized into a scarce resource to enter the conventional economy. Within the conception of ambient economics, ambient goods such as data take a scarce form through the assigning of legal form via mediating institutions. It is at this moment, where data is assigned a legal form, that surveillant actors are at their most economically vulnerable. The legal/regulatory decisions do not only limit how they may use the data they have participated in synthesizing to fulfill their desires but also how they may leverage it as an economic instrument, and the value of their use rights.

Although legislation such as the EU GDPR or the Canadian Digital Charter may be generally communicated as privacy legislation, they can also be framed as economic legislation that governs the form data, or the use rights to it, may take in the conventional economy. As noted by Scassa (2020) data privacy legislation can also be interpreted as legislation to legitimize the collection and processing of user data for economic reasons, by providing accepted rules. For example, in the EU, recent legislation such as the Data Act or Digital Market Act have provisions that provide individuals, and businesses, rights to access and use data produced via their interactions with the surveillant company. This reduces the use value of the data to the company, as it removes the company's exclusivity to the data. The California Consumer Privacy Act (CCPA) enshrines extremely specific limitations on the sale of access to personal information, including the necessity for a clear path for the data subject to opt-out (California Consumer Privacy Act, 2018). The EU GDPR requires that data only be processed for the

reasons consent was originally given, again restricting the ability to leverage the sale of data access as a commodity (General Data Protection Regulation, 2016).

One of the most interesting ways the ambient economic function of these mediating regulations can be observed is through a change made to Canada's Bill C-27. Bill C-27 is part of Canada's Digital Charter initiative, proposing changes to update the *Consumer Privacy and Protections Act* (Canada, 2023). The Bill positions itself as putting Canadians' Privacy at the forefront, providing a balance between Canadians' rights to privacy, **and the legitimate interests of businesses for the collection and processing of data** (Scassa 2022). The Bill presents robust rules for informed consent when collecting personal data, and provides avenues for citizens to have their collected personal data erased. However, the focus of this bill is on personal information, or personal data. Personal information, within the scope of the bill, is defined as "information about an identifiable individual." (Parliament of Canada, 2023). Therefore, provisions such as the requirements for consent before collecting (what I refer to as synthesizing data) apply only when the data collected can be linked back to the individual it is collected from. Further, there are provisions in the Bill which allow companies to anonymize the data they collect (synthesize in my conception), rendering the said data outside of the scope of the legislation (Scassa, 2022). I suggest this function also produces the practical disentanglement of the data from the surveilled subject, as it is no longer considered 'personal' and therefore no longer subject to the rights claims/constraints placed on personal data that restrict it from entering the conventional economy itself. The bar for anonymizing is operationally high, as it requires that there be no future ability to use the data to identify its source (Scassa, 2022). However, Scassa (2022) notes that due to industry lobbying a second classification was added: de-identification. De-identification refers to removing direct identity markers from the data (Scassa, 2022). Notably, this is a downgrade from previous Canadian privacy legislation, wherein data was not considered de-identified unless indirect identity markers were removed as well (Scassa, 2022). Within the scope of the Bill, as long as individual identifiers are removed, a company may collect, and use a subject's data without their knowledge or consent (Scassa, 2022).

In this instance, a simple change results in a completely different outcome for the surveilling actors, providing them with greater ability to collect data which they can stake exclusive access to, and transform into a commodity if they desire. Further, it positions data synthesized via surveillance of a human subject that does not contain information about an identifiable individual as outside the privy of the regulation, legitimizing the data's severing from the surveilled subject for commoditization. Thus, the surveilling actor may be perceived as having an ownership –or at least de facto ownership-like– right to the data, allowing them to introduce the data itself into the traditional economy as an economic instrument and subject.

Positioning legislation and other mediating processes that give data its scarce, and legal, form –so that it may enter the conventional economy– as economic actions is vital to meaningfully understand and control the economic and social effects of data and AI development. By understanding, and providing language for the process through which data becomes an economic instrument, it can be better studied and regulated. Further, I suggest that the concepts developed in this text can aid in understanding the economic motivations of surveillant actors regarding the development of data legislation that may not otherwise be apparent. This paper raises questions about how we construct mediating institutions, in order to extract the maximum possible value from data. Data gains greater use value through the discovery of new potential uses of the data. The use value is not finite, and is in constant flux. Allowing greater access to a given set of data, to be experimented on, or combined with new data, may create greater use value, and produce more long-term economic benefit. However, doing so comes at the cost of lowering the relative use value of the data to an actor who wishes to make the data exclusive. Further, it sacrifices the economic benefits of creating an exchange value for the data.

I propose that if it is the goal of conventional scarcity economics to provide and produce structures that allow society to best utilize and distribute our finite resources; then it should be the ambition of ambient economics to create structures and systems that allow society to realize the maximum use value of ambient goods, while simultaneously adhering to an respecting the external constraints of society. The most important question is not "How do we value data" but "How do we make data as valuable and productive for society as possible". It is my position that when examined through the lens of ambient economics, 'privacy' legislation such as Bill C-27 artificially restricts the optimization of the value of data co-produced by social actors, by legitimizing its capture as a scarce resource, and its control by a singular market actor.

CONCLUSION

What is presented in this paper is not a radical new form of economic thinking. Instead, it is a proposal to re-situate data as a subject within our current understanding of markets and property. The concept of the ambient economy is a tool that fits within established economic knowledge. However, applying the concept to data illuminates why traditional economic methods struggle to adequately address vital questions regarding data as a subject of economic inquiry, such as its value.

It is argued that to take on exchange value within the market economy, data must be transposed into a scarce form through the assignment of exclusionary rights, or de-facto scarcity due to a lack of access ability by other actors. The concept of the ambient economy aids in understanding why data's journey to gaining scarce form is different from other non-scarce goods, such as a story, which are made scarce by mechanisms such as IP law. The difference is born from the unique situation of data production via surveillance, and the social rights claims that the action of

surveillance invokes. These rights claims complicate the disentanglement process which is necessary for a non-scarce good to be given a legal form that will allow it to become an economic subject in the market. As I explore, this often leads to data not being able to enter the market as an economic subject, instead, it is often a completely different product, use rights, that is traded or exchanged for value.

The final section of this text explores a practical implementation of these theories. It is argued that when examined through the lens of the ambient economy, privacy legislation such as Bill C-27 is acting as de facto economic legislation. The argument is not against privacy legislation, as preserving the right to privacy is paramount. However, this form of legislation is actively shaping the ambient economic mechanisms through which data can enter the conventional scarcity economy, assigning scarcity and other economic rights to surveillance actors. Such decisions must be analyzed as actions with long-term economic consequences regarding how data as a resource is managed.

BIBLIOGRAPHY

- Allmer, T. (2012). Towards a critical theory of surveillance in informational capitalism.
- Alston, L. J., & Mueller, B. (2008). *Property rights and the state*. In *Handbook of new institutional economics* (pp. 573-590). Springer, Berlin, Heidelberg.
- Andrejevic, M. B. (2011). Surveillance and alienation in the online economy. *Surveillance & Society*, 8(3), 278-287.
- Andrejevic, M. (2019). Automating surveillance. *Surveillance & society*, 17(1/2), 7-13.
- Andrews, L. B. (2012). I know who you are and I saw what you did: Social networks and the death of privacy. Simon and Schuster.
- Arrow, K. J. (1973). *Information and economic behavior* (Vol. 28). Stockholm: Federation of Swedish Industries.
- Arrow, K. J. (1978). Uncertainty and the welfare economics of medical care. In *Uncertainty in economics* (pp. 345-375). Academic Press.
- Batra, N. D. (2007). *Digital Freedom: How Much Can You Handle?*. Rowman & Littlefield Publishers.
- Boyle, J. (2008). *The public domain: Enclosing the commons of the mind*. New Haven, Conn.; London: Yale University Press, 2008.
- Buckland, M. (2017). *Information and society*. Mit Press.
- Callon, M. (1998). *The laws of the markets*. Oxford: Blackwell.
- California Consumer Privacy Act. (2018). California Civil Code § 1798.100 et seq. <https://coppa.ca.gov/regulations/pdf/coppa_act.pdf>
- Canada. (2023b, March 13). *Consumer Privacy Protection Act*. Language selection - Innovation, Science and Economic Development Canada Main Site / Sélection de la langue - Site principal d'Innovation, Sciences et Développement économique Canada. Retrieved April 17, 2023, from <https://ised-isde.canada.ca/site/innovation-better-canada/en/consumer-privacy-protection-act>
- Couldry, N., & Mejias, U. A. (2020). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

- CPA Canada. 2021. What Is Your Data Worth? Insights for CPAs. Chartered Professional Accountants Canada.
- Duch-Brown, N., Martens, B., & Mueller-Langer, F. (2017). The economics of ownership, access and trade in digital data
- The Economist. (2017, May 6). *The world's most valuable resource is no longer oil, but Data*.
https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18798097116&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gclid=Cj0KCQjw7JOpBhCfARIsAL3bobf3eTnrAvU_7tM3hms5HMdK1pLMYtL_Sla83GDa8lMyLCrVR2C_4M8aAuBOEALw_wcB&gclsrc=aw.ds
- Eliot, David and David Murakami Wood. 2021. "Minding the FLoCs: Google's Marketing Moves, AI, Privacy and the Data Commons." *Centre for International Governance Innovation*. Retrieved July 5, 2021
 (<https://www.cigionline.org/articles/minding-flocs-googles-marketing-moves-ai-privacy-and-data-commons/>).
- Eliot, D. (2022). *Towards a Theory of Ambient Capitalism: An Inquiry into How Data Creates Value* (Masters dissertation).
- Fuchs, C. (2010). "Web 2.0, Prosumption, and Surveillance." *Surveillance & Society* 8(3):288–309. doi: 10.24908/ss.v8i3.4165.
- Fuchs, C. (2012). Google capitalism. *tripleC: Open Access Journal for a Global Sustainable Information Society*, 10(1), 42-48.
- Fuchs, C. (2019). *Rereading Marx in the age of digital capitalism*. Pluto Press.
- General Data Protection Regulation. (2016). Official Journal of the European Union, L119/1-L119/88
 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.001.01.ENG&toc=OJ%3AL%3A2016%3A119%3ATOC>
- Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), "Raw Data" Is an Oxymoron. The MIT Press. <https://doi.org/10.7551/mitpress/9302.003.0002>

- Google Safety . (n.d.). *AD controls & personalization on google - google safety center*. Ad Controls & Personalization on Google - Google Safety Center.
<https://safety.google/privacy/ads-and-data/#:~:text=information%20to%20anyone.-,We%20never%20sell%20your%20personal,anyone%2C%20including%20for%20ads%20purposes.>
- Jessop, B. (2007). *Knowledge as a fictitious commodity: Insights and limits of a Polanyian perspective*. In S. Charusheela & E. M. Wood (Eds.), *Reading Karl Polanyi for the twenty-first century* (pp. 115-133). Palgrave Macmillan.
- Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.
- Litman, J. (1999). Information privacy/information property. *Stan. L. Rev.*, 52, 1283.
- Mandel, G. N. (2006). Patently non-obvious: Empirical demonstration that the hindsight bias renders patent decisions irrational. *Ohio St. LJ*, 67, 1391.
- Mankiw, N. G., & Taylor, M. P. (2014). *Economics*. South-Western Cengage Learning.
- Mansell, R. (Ed.). (2007). *The Oxford handbook of information and communication technologies*. Oxford Handbooks.
- Marx, K. (2019). *Capital: Volume One*. Courier Dover Publications.
- Marx, G. T. (2002). What's New About the "New Surveillance"? Classifying for Change and Continuity. *Surveillance & Society*, 1(1), 9-29.
- Parliament of Canada (2023). *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*. Retrieved from the Parliament of Canada Website:
<https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big data & society*, 6(1), 2053951718820549.
- Samuelson, P. (1999). Privacy as intellectual property. *Stan. L. Rev.*, 52, 1125.
- Scassa, Teresa, Data Ownership (September 4, 2018). CIGI Papers No. 187, Ottawa Faculty of Law Working Paper No. 2018-26, Available at SSRN: <https://ssrn.com/abstract=3251542> or <http://dx.doi.org/10.2139/ssrn.3251542>

- Scassa, T (2020, November 12) The History of Privacy and the Future of AI with Dr. Teresa Scassa [Video]. Youtube. <https://www.youtube.com/watch?v=MXGsRk8CL1o>
- Schneier, B. (2015). *Data and Goliath: The hidden battles to collect your data and control your world*. WW Norton & Company.
- Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.
- Statistics Canada. (2019, July 10). The value of data in Canada: Experimental estimates. Retrieved September 26, 2023, from [https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm]
- Stonier, T. (2012). *Information and meaning: An evolutionary perspective*. Springer Science & Business Media.
- Sumpter, D. (2018). *Outnumbered: From Facebook and Google to Fake News and Filter-bubbles—the algorithms that control our lives* (Vol. 36). Bloomsbury Publishing.
- Tang, C. (2021). *Data Capital: How Data Is Reinventing Capital for Globalization*. Cham, Switzerland: Springer.
- World Trade Organization. (n.d.). *WTO Rules and Environmental Policies: Key GATT Disciplines*. Retrieved March 29, 2022, from https://www.wto.org/english/tratop_e/envir_e/envt_rules_gatt_e.htm
- Zhou, Z. H. (2021). *Machine learning*. Springer Nature.
- Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology*, 58(4), 479-493.
- Zion, I. B. (2021, January 18). *Israel Trades pfizer vaccine doses for medical data*. PBS. <https://www.pbs.org/newshour/science/israel-trades-pfizer-vaccine-doses-for-medical-data>
- Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism : the Fight for a Human Future at the New Frontier of Power* . First edition. New York, NY: PublicAffairs.
- Zuboff, Shoshana. 2015. “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization.” *Journal of Information Technology* 30(1):75–89. doi: 10.1057/jit.2015.5.

