FUTURE ECONOMIES RESEARCH AND POLICY PAPER #7

January 2020

# Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership

**Stuart Mills**

**Manchester Metropolitan University**

**Future Economies UCRKE**

**About the Author**

Stuart Mill's research considers user responses to the commercial use of behavioural economics. He has previously written about the political application of behavioural economics, contemporary political economy and technology.

## About Future Economies

Future Economies, a university centre for research and knowledge exchange based at Manchester Metropolitan University, brings together academics from a wide range of disciplinary backgrounds, alongside policy and business practitioners, to conduct research into local, national and global economic challenges, ranging from Brexit, financial crisis, devolution and local industrial strategies to mega-sporting events and trade governance. Future Economies has a particular expertise in political economy and behavioural economics, and also encompasses Future Economies Analytics, the Centre for Policy Modelling and the Sports Policy Unit.

***All views expressed in this paper are those of the author, and are not necessarily shared by Future Economies or Manchester Metropolitan University.***

# Table of Contents

## Abstract

Who owns your data? And why do they? In this article, I consider various stakeholder claims to data ownership and the value generated by data, through a political economy lens. Following a data value framework established by the Open Data Institute, I first consider how data generates value from the point of creation, how data as a resource imbue various stewardship obligations onto data controllers, and finally how – given competing interests – decision-making authority is apportioned across stakeholders. This framework is then applied to three emerging models of data ownership: Laissez Faire, Data Trusts and Data Commons. The structural qualities of each model are revealed by an in-depth critique, before a visualisation of the data flows between stakeholders is offered. Finally, I compare these models across categorical issues that emerge from this analysis, considering how each model tackles issues such as incentives, competition, innovation and feasibility.

# Introduction

In 2017, the Economist featured an article headlined, "The world's most valuable resource is no longer oil, but data" (Economist, 2017). To many, this will not be surprising; at the time of writing, Forbes reports the five most valuable companies in the world are all in the technology and data sector (Forbes, 2019).[1]

The value of data need not be measured financially, however. Recent data incidents such as the Cambridge Analytica scandal demonstrate the power data have over elections (Confessore, 2018; Zittrain, 2013), over our behaviour (Beer, 2017), and over our regulatory capacities (Yeung, 2017).

Regulation is a particularly interesting question within our current data environment, for two reasons. Firstly, there is a data access question which emerges when vital data are privately held. For example, Young et al. (2019) and Yeung (2017) argue governments may be ill-equipped to regulate big data industries without access to big datasets. Hall and Pesenti (2017) have raised a similar issue in their report on the AI industry for the UK's Departments of Business and Culture. They suggest that without access to data, neither the public nor private sector will be able to fully exploit the benefits of AI.

Secondly, there is a regulatory question of how to tackle market challenges presented by ever-more monopolistic data firms (Srnicek, 2016). Contender for the Democratic party nomination for President Elizabeth Warren has made breaking up large technology firms using antitrust legislation a cornerstone of her presidential program (Warren, 2019). By contrast, Facebook CEO Mark Zuckerberg has recognised the need for regulation but has argued a platform such as Facebook should not be broken up (Isaac, 2019). The competing arguments over breaking up big tech is explained by Srnicek (2016), who argues the effectiveness of these platforms is their size and ability to lever, "network effects" (Srnicek, 2016: 47). Srnicek suggests in this environment few easy regulatory responses exist: presently, they argue data firms are likely to become enclosed and congruent as these firms seek to control data flows, fragmenting digital spaces such as the internet and creating high barriers of entry, but breaking up these firms

[1] These are Apple, Google, Microsoft, Amazon and Facebook.

undermines the network effects that make them both desirable services and profitable businesses.

Given a regulatory response is required, but an antitrust approach may undermine the functionality of data firms, a literature surrounding our present and possible alternative data ownership models has emerged. Much of this literature has either been technical (ODI, 2019; Young et al., 2019; Grossman et al., 2016) or legally focused (O'Hara, 2019; Hall and Pesenti, 2017), exploring questions such as the data infrastructure (ODI, 2019) required for alternative models, or how to build a legal framework for these models (O'Hara, 2019; Yeung, 2017).

This is curious, as the question of ownership (data or otherwise) is also a question for political economy. This paper adopts this perspective, and proceeds, initially, to address two questions which emerge:

- What are data, from the perspective of how is it produced, and which actors are involved in its production?
- What are the claims to ownership with emerge from the production of data?

These questions are intimately linked. Mazzucato (2018), for example, argues:

"[T]hat a large part of the technology and necessary data was created by all of us, and should belong to all of us. The underlying infrastructure that all these companies rely on was created collectively (via the tax dollars that built the internet), and it also feeds off network effects that are produced collectively. There is indeed no reason why the public's data should not be owned by a public repository that sells data to the tech giants, rather than vice versa"

However, counter-arguments oppose the origin of data which underpins Mazzucato's argument. Gitelman (2013), for example, argues that data only become data when it is conceived as such. We all have an age, a gender, a nationality and so on. But these facets of our being do not become data until they are noted and recorded. As Srnicek (2016) – invoking Marx – puts it, "simply put, we should consider data the be the raw material that must be extracted, and the activities of users to be the natural source of this raw material" (Srnicek, 2016: 40). Despite the potential confusion this dispute over the origin of data raises, a parallel in political economy exists, namely the tension between labour and capital described in Hodgskin's (1825) work.

Finally, separate from production and ownership, but necessary for both, is the transactional nature of data. O'Hara (2019) argues out modern understanding of data has adopted the ideological individualism of neoliberalism. In more explicit terms, Yeung (2017) argues data are falsely viewed as an individual resource to be transacted. They suggest technical advances allows tech firms to subvert this transaction, capturing data about those who have not consented, and undermining this 'transactional' relationship.

All these arguments are vital considerations for new models of data ownership which existing literature has yet to fully synthesis. Through a political economy lens, I explore these questions and appraise the emerging, alternative models to data ownership.

The structure of this article is as follows: Section 2 investigates what data are and the relationship between claims to value generated from data, and data ownership, using a value framework established by the Open Data Institute. Section 3 uses this framework to analyse three possible data ownership models: laissez faire, data trusts, and data commons. Section 4 compares these models by considering the incentive, competitive and innovative qualities of these models, as well as the feasibility of implementation. Section 5 concludes.

## A Note on Terminology

Throughout this article, various terminology is used to describe actors within the data production process which may differ from previous work in this area. For example, O'Hara (2019) uses the term, "data controller," to describe an actor who controls data. In the context O'Hara uses this phrase, it is perfectly adequate. Yet, insofar as a study of data ownership is concerned, an emphasis on the actor who collects data is more important. This actor may also be a data controller, but this term does not adequately consider the originator role which is of interest.

Likewise, the term data subject is often used to describe a person about whom data are collected. Once more, this skews the focus away from this actor's role in the data production process.

Throughout this article, I will use the term, "data collector," to describe an actor who collects data (in the language used above, who conceives of observations as data) and the term, "data generator," to describe an actor whose actions are conceivable as data, and about whom data describe. The term, "data service," is used to describe an actor who uses data, though this actor

does not have to be the data collector. Where additional terminology is used, a sufficient definition will be provided.

# Data as a Valuable Resource

## What are Data?

Various authors draw a distinction between information and data (Determann, 2018; Scassa, 2018; Kitchin, 2014; Rees, 2013). Gitelman (2013) argues information become data only when they are conceived and recorded as such. Kitchin (2014) offers a similar argument, recognising what Scassa (2018) calls the, "non-neutrality," (Scassa, 2018: 3) of data. This principle supposes data emerge from the intentional collection of information. This intentionality is also implicit in Gitelman's (2013) argument: for one to conceive of information as data and subsequently collect them, one presumably acts with intention.

Rees (2013) captures this notion too, but from a somewhat different perspective. Broadly, they seem to agree with the conception or non-neutrality arguments of Gitelman (2013), Kitchin (2014) and Scassa (2018), but further argue that information become data only through combination with other information (also see Srnicek, 2016). Rees (2013) gives the example of names and addresses. Separately, these are just pieces of information. Combined, however, they become data, as these data now enable, say, billing of customers.

The combination postulate is flawed for two reasons. Firstly, intentionality can supersede combination. For example, a list of names become data if one wants to know the popularity of a given name within the sample. Presumably this is a sufficient transformation to render this information data and is completed without the need to combine it with any other.

Secondly, Kitchin (2014) argues data take three forms: representative, implied and derived. Representative data are that which can be observed and measured (Scassa, 2018), following – presumably – first being conceived as data (Gitelman, 2013). The information Rees (2013) gives to justify their combination argument (names, addresses), then, instead resemble representative data. Implied data are data which representative data imply. For example, from a sample of names, the name, 'John,' is likely implied as a popular name due to its frequency, despite no representative data that directly capture opinions about the name John. Finally,

derived data are data produced from other data. Rees' (2013) example of combining names and addresses to produce data, then, appears to be an example of derived data.

Determann (2018) argues the approach to data creation outlined by Gitelman (2013) and Kitchin (2014), and particularly Kitchin's derived data argument, is problematic for the notion of data ownership. Determann (2018) argues that if data are merely the process of conceiving of information as data, the amount of data seems to become infinite and any claims to ownership highly tenuous.

There is some merit to this argument. Scassa (2018) cites Marr (2018) and Schlosser (2018), both of whom argue the comparison of data to oil is inaccurate because data don't run out. Indeed, Scassa (2018) states, "data are an infinitely renewable resource" (Scassa, 2018: 1). Ownership is a function of scarcity (Bastani, 2019). If data are not scarce, but rather abundant, the notion of data ownership seems superfluous (Determann, 2018). This argument, however, is one bound in a theoretically economic conception of data rather than a material one.

Firstly, while the marginal costs of reproducing data may be (or close to) zero (Marr, 2018; Scassa, 2018; Schlosser, 2018), the initial costs of collecting data, which is to say the cost of conception, is not zero (Srnicek, 2016). Further, existing and possible data ownership structures such as data monopolies (Lawrence and Laybourn-Langton, 2018; Scassa, 2018) may produce manufactured scarcity. [2] Determann's (2018) argument may hold some validity (and increasingly so) on the question of derived data and surrounding technologies such as big data, where the cost of conception is the marginal cost. This question, however, is beyond the scope of this paper.

Secondly, the concept of non-neutrality reveals a scarce quality of data. Namely, non-neutral data are also data that are subject- and context-specific. While data may be infinitely renewable in terms of use and reproduction, this is a function of information. For example, no matter how many times one's age is recorded and reproduced, that information is not destroyed. Yet age information only become data when applied within a context or towards a subject. Social media data, for example, are only data when applied to the individual about whom it pertains, or within the population they were created, while atmospheric temperature data are only an observation until applied within the environment they are observed. This is the substantial

[2] For more, see Bastani's (2019) discussion on biometric data and the use of copyright law.

consequence of non-neutrality and reveals a twist (or perhaps nuance) of Rees' (2013) combination argument: data are produced through the combination of information and context, the latter captured by the concept of non-neutrality.

## A Framework of Data Value

The ODI (2019) argue three characteristics of data (from conception, through handling and ultimately application) describe the source of their value. These characteristics are: creation, stewardship, and decision making.

### Creation

If data are the combination of information and context, the claims to either of these constituents by separate parties highlights a tension at the heart of data ownership. These constituents, too, reveal some characteristics of data, namely that data are neither expendable (Marr, 2018; Scassa, 2018; Schlosser, 2018) nor fungible (Scassa, 2018; Kitchin, 2014; Gitelman, 2013).

A parallel to this data ownership tension can be seen in the classic political economy argument analysed in Hodgskin's (1825) Labour Defended Against the Claims of Capital. Hodgskin argues that the major role of capital in the production process is to provide labour with the resources necessary to be productive. For example, a pin maker (à la Smith) cannot perform their craft without materials and tools, which capital provides. However, Hodgskin argues that at any point within the productive process (culminating in the extraction of raw materials, which one might claim are morally part of a collective common), it is labour and not capital which control the means of production. Thus, he argues, production could continue without capital, but not without labour.

Regardless of one's affinity towards Hodgskin's argument, it is possible to present a similar dichotomy on the question of data creation.

- Argument 1 – The "Capital" Claim to Data
  There is an unlimited amount of data, and the marginal cost of data is zero (Scassa, 2018). However, there is a cost of conception that is incurred during the data collection process and only by those that conceive of and collect data (Gitelman, 2013). Therefore, data collectors have legitimate claim to the data they collect because of the cost of

collection, and legitimate claim to the value generated by data because without their efforts, no value would be created.

- Argument 2 – The "Labour" Claim to Data
  Data are not an expendable resource, and so once data are collected, the marginal cost of using data is zero, with repeated data usage driving the cost of collection to zero (Determann, 2018). Yet data are not a fungible resource. Specific data can only tell us about the specific subject. As such, data have no value unless applied in conjunction with the subject whom the datum pertains to (Kitchin, 2014; Rees, 2013). Therefore, the source of data's value are data generators, and thus they have a legitimate claim to both value and ownership (Mazzucato, 2018).

Again, regardless of one's sympathies with either argument, there is clearly a familiar tension between actors in the production of data, and actors in classical political economy. Just as in the latter this tension spurs on debates regarding models of ownership within economies, so too should this tension spur debates about models of ownership within the data economy.

## Stewardship

Stewardship is the central consideration in the ODI's (2019) work on data trusts. However, their definition of a data steward reveals applications of the concept extend beyond this model of data ownership: "[A] steward of data… can decide who has access [to data], under what conditions, and to whose benefit" (ODI, 2019:6, emphasis added). Questions of access, pre-conditions and who receives value are all intimately woven into the question of data ownership.

For example, while not directly comparable to the treatment of data (which continues as a developing field. See Coyle and Nguyen (2019) for more), the International Accounting Standards Board (IASB) Conceptual Framework for Financial Reporting (2018) defines an asset as, "a resource controlled by the entity as a result of past events and from which future economic benefits are expected to flow to the entity" (IASB, 2018:2). Within this accounting grey area, then, until such a time as further clarification is provided, one might argue data stewardship confers the benefits – if not the reality – of data ownership onto whomever fulfils the criteria for access (O'Hara, 2019).

This is but one challenge that faces the concept of stewardship. Another, which the ODI (2019) and O'Hara (2019) acknowledge, is that the potential structure of data ownership models. The ODI and O'Hara argue – specifically regarding data trusts – that the exact structure of this ownership model is still developing, and further, flexibility in structure may be desirable to promote the use of data trusts. However, the structure of data ownership models (data trusts or otherwise) have important consequences on the obligations of the data steward, and thus the flow of data's value.

A clear example of this emerges from the British Academy and Royal Society's (2017) call for human flourishment to guide data governance. While a noble goal, such purposeful vacuity (O'Hara, 2019) defers answering structural questions about data ownership. Further, human flourishment becomes a point of false equivalency between different data ownership structures which distribute the value of data in different ways. Is fiduciary duty and maximising value human flourishment, as a free market perspective might suggest? Or is human flourishment about protecting privacy and promoting democratic decision making, as a human rights advocate might argue?

Stewardship, then, exists within an ownership structure. Thus, the specific purpose of the data steward must reflect the ambitions of those who build data structures, as well as any tension that may result from additional stakeholders. This is a political economy question.

## Decision Making

Decision making is the final layer of value creation using data and emerges from the creation and stewardship layers. Decision making is also intimately tied to data ownership; as above, the use of data falls within the remit of the data steward, whose role is defined by the data ownership structure.

The relation between decision making and stewardship seems so close one might be tempted to not draw a firm distinction between the two, though this seems inappropriate. For instance, decision making should be recognised as taking two forms, firstly decision making over data, and secondly decision making over the governance of the data owner (Srnicek, 2016).

As an example, a data steward might give a data service access to data. This data service will be bound in their decision making by the terms of access (which may be very loose or highly

rigorous), yet within these confines will be free to make any decision they wish. But decision making regarding the terms of access is separate from the decision of who gets access based on these terms. This returns to a structural question of data ownership.

There are wider, more moralistic questions regarding decision making too. Acquisti, Brandimarte and Loewenstein (2015) argue social media users often don't understand the terms they agree to when handing over data, highlighting an asymmetry of knowledge between actors (Lanzing, 2018; Yeung, 2017; Helbing et al., 2017). Allowing, tackling or entrenching such asymmetries are further decisions which flow from the data ownership structure.

## New Models of Data Ownership

So far in this article I have presented two arguments. Firstly, there is a conflict between actors – revealed by a political economy analysis – in the production process of data because of the unique qualities of data as a resource, namely they are neither expendable nor fungible. Secondly, that this conflict continues to shape the flows of value that data create, revealing nuances such as stewardship and decision making which reflect the ownership structure of data.

Passing reference has been made to the three data ownership models considered by this article. These are laissez faire ownership, data trusts and data commons. In this section I will unpack each of these models, highlighting the differences in data ownership structure, before considering each model through the lens of the creation, stewardship and decision-making framework used above.

I should also note that the names I have given each of these models may be non-standard and may create confusion without proper explanation. For example, a laissez faire ownership model contains the feature of data monopolies, which some might contend is a more appropriate description of this ownership model. In choosing a naming structure, however, I have chosen to focus on how data as a resource are treated within each model. Continuing the example, then, under laissez faire, data are seen as an individual resource which individuals are free to transact with. Taking this approach, the term, 'data monopoly,' seems unrepresentative.

## Laissez Faire

Laissez faire data ownership is best described as the model most commonly used today by private data service companies. The language of laissez faire, free data and free market data all borrow from the neoliberal turn in data conceptualisation outlined by O'Hara (2019). Likewise, in their discussion of privacy in the big data age, Yeung (2017) outlines a pervasive attitude towards data consisting of barter, building on the work of van Dijck (2014), where private individuals (data generators) provide their data on a pseudo-transactional basis in exchange for, 'free,' services.

In this sense, the data ownership question is both obscured and simplified; rather than ownership deriving from value creation, the laissez faire model sees data ownership conferred onto data collectors via a traditional transaction approach.

Following from this, data generators are considered to receive value through the provision of free or pro-convenience services, with consent assumed based on a generator's decision to give data over to the data collector (Yeung, 2017). Thus, any additional value these collectors can extract from these data is solely theirs, as is any decision-making authority once consent have been received.

This is not to say that the laissez faire model does not impose some constraint onto data collectors. Instead, these constraints are shaped more ideologically (van Dijck, 2014), rather than through moral of ethical compunction. Broadly, three constraints exist: external regulation, fiduciary responsibility, and market forces.

External regulation includes legislation such as the European Parliament's (2016) GDPR data protection regulation. Regulation may also exist governing data sharing (for example, medical records), transparency (article 15 of GDPR requires data collectors provide a copy of data to those from whom they collect. See O'Hara (2019) for more), civil protections (limiting the amount of data that can be collected on minors or adding additional protections) and rights protections (for example, article 12 of the UN Universal Declaration of Human Rights (1948) enshrines privacy as a right).

Fiduciary duty may also be described as a form of regulation but is better conceptualised as the data collector's obligation to maximise the return to shareholders, or the benefit to chosen beneficiaries.

Finally, market forces include concepts such as competition, innovation and reputational risk, all of which can change the data-driven strategy of the data collector, and thus their decision making. However, the extent to which real-world data collectors are subject to market forces is disputable. For instance, under the laissez faire model, all new entrants to the market must build their own datasets, creating high barriers to entry (Hall and Pesenti, 2017; LeCun, 2016; Srnicek, 2016). Furthermore, as Hall and Pesenti (2017) argue, as data become big data and datasets grow exponentially, data monopolies may develop.

**FIGURE 1 – Data Flows in a Laissez Faire Model**[3]

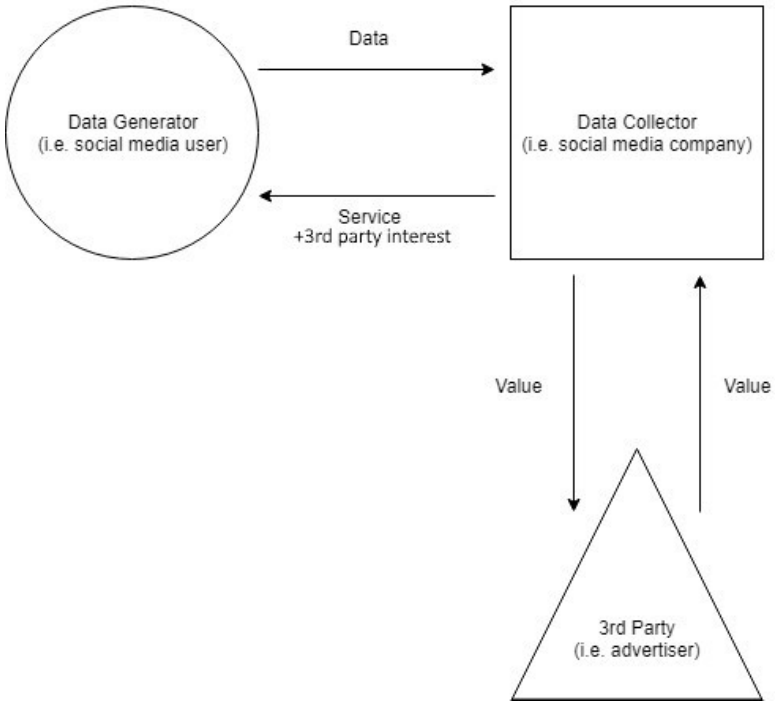



Figure 1 illustrates the data flow within the laissez faire model. Data generators and collectors engage in a transaction of data for services, while the collector engages in an exchange of value with a 3rd party, whose interest is subsequently passed towards the generator.

## Data Trusts

The ODI (2019) define a data trust as, "a legal structure that provides independent stewardship of data," (ODI, 2019: 6), however this definition is potentially problematic when juxtaposed with that of O'Hara's (2019) work. O'Hara argues a data trust is not a trust in a legal sense.

[3] Author's own.

Rather, it is a trust in the literal sense of being a group of partners who all trust each other. Thus, it seems questionable to define a data trust around a legal structure.[4] However, as the ODI (2019) note from their discussion of several pilot data trusts, the circumstances and requirements of each pilot created different data trusts, and so a legal definition of a data trust may instead be considered an ambition rather than an ironclad definition.

Equally, designing data trusts to suit different contexts and actors (as O'Hara (2019) encourages) undermines a second aspect of the definition given by the ODI – namely, how can the independence of the data trust be guaranteed? If, as we might accept, that the legal structure of a data trust should be adaptable, so too will the obligations of the trustees who manage the trust. Guaranteed independence, then, is perhaps a theoretical desirability but a practical difficulty.

Alternatively, Hall and Pesenti (2017) define a data trust as, "proven and trusted frameworks and agreements… to ensure exchanges are secure and mutually beneficial," (Hall and Pesenti, 2017: 46). They elaborate further, "trusts are not a legal entity or institution, but rather a set of relationships underpinned by a repeatable framework" (Hall and Pesenti, 2017: 46).

Immediately, a reaffirmation of O'Hara's (2019) argument that data trusts are not legal entities can be seen. Further, Hall and Pesenti choose to elevate the role of data trusts (such as data sharing) over concepts such as stewardship advanced by the ODI (2019), championing repeatability such that data trusts may become a vehicle for data sharing. This is not surprising given Hall and Pesenti advocate data trusts to encourage the development of the UK AI industry, while the ODI (2019) are much more concerned with responsible and meaningful data sharing.

Regardless of this definitional uncertainty, the purpose of this discussion is twofold. Firstly, to demonstrate the developing nature of this work. Secondly, to highlight the common themes throughout this work, namely the presence of a third-party trust which occupies the role of a data steward on behalf of various stakeholders.

It is this second point which is of most immediate interest. The existence of a data trust reimagines data not as an individual resource described by Yeung (2017), but as a pooled resource which is greater than the sum of its parts. From a material perspective, this very much

[4] Beyond considering the contractual relationships between actors as a legal structure, but such an approach renders almost all data bound in some legal structure, and thus potentially a trust.

rings true, given the potential advantages of big data and AI (Hall and Pesenti, 2017). However, viewing data as a pooled resource raises once more the question of ownership and value.

O'Hara (2019) argues the trust should own the data but does so only for the benefit of its members. Unfortunately, this raises more questions than answers. For example, the ODI (2019) recognise the funding problem which emerges with a data trust – is the trust funded by its members, which compromises the independence of the trust, or does the trust operate for-profit, which may undermine its purpose, or is the trust externally funded? Furthermore, returning to the claims of ownership by data generators and data collectors, for whom should a data trust act if both groups claim to have a legitimate claim to ownership?

These questions are tackled in the following constructions of data trusts.

## Collector Centric Data Trust

The collector centric data trust is most typical of the existing work on data trusts.[5] Under this model, data collectors continue to amass data from generators as they do under the laissez faire model. Thus, data only resemble a pooled resource from the perspective of data collectors – for data generators, this model of ownership appears very similar to the laissez faire model, and so data may still be conceived as an individual resource.

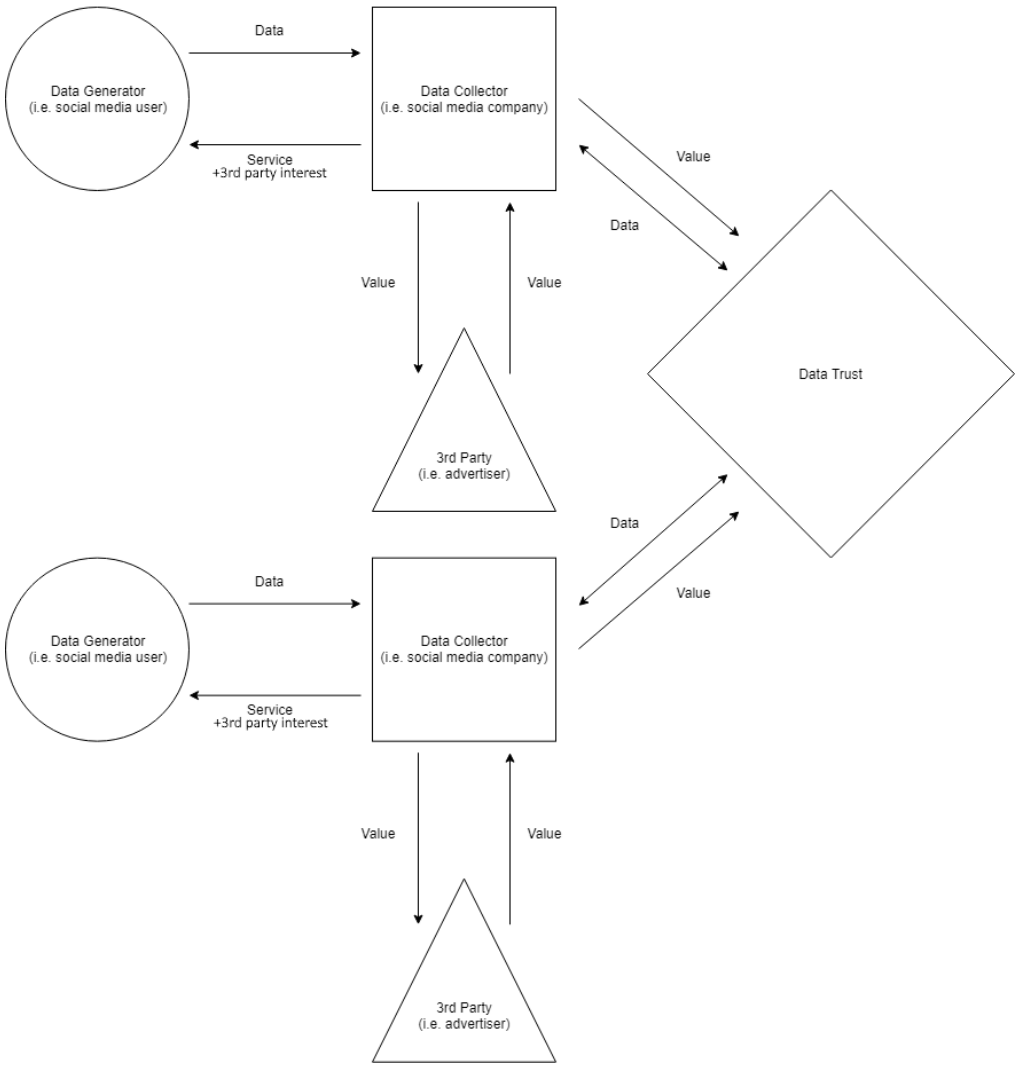
Collectors then pool their data into a data trust, each sharing in the advantage of larger datasets. Some of the value collectors generate via 3rd parties is returned to the trust to cover funding costs.

In a collector centric data trust, the trust's stewardship responsibilities are to protect its members, providing relatively low-cost access to data. Furthermore, the trust may also operate commercially, selling access to data at a premium to market entrants and distributing this revenue amongst members, or denying access to new entrants to protect market incumbents. In

[5] For example, see Collaborative Data Trusts (CDT) put forth by Young et al. (2019). The collector centric data trust and CDTs differ as Young et al. (2019) emphasis public-private data trusts, similar to the suggestion of Hall and Pesenti (2017), while the collector centric data trust is more generalised to emphasis data flow.

this sense, a collector centric data trust resembles monopolistic market coordination, and may stymie competition, contrary to the arguments of Furman (2019) and Hall and Pesenti (2017).[6]

**FIGURE 2 – Data Flows in a Collector Centric Data Trust[7]**




[6] Hall and Pesenti (2017) imply a larger role for government, possibly as a member of the trust, possibly as the trustee of the trust, possibly as the funder of the trust. Regardless, there may be a role for government which reduces the potential downsides of this type of trust for consumers.

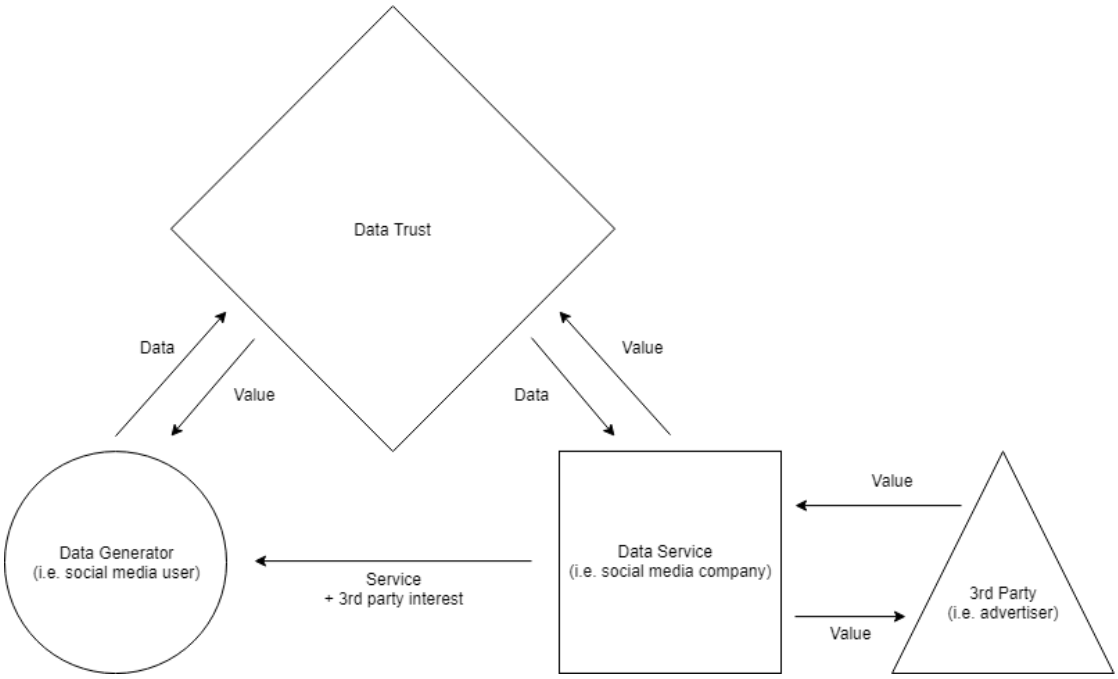
[7] Author's own.

## Data Centric Data Trust

A data centric data trust is very much a reversal of the collector centric data trust. Here, data are sincerely treated as a pooled resource, with data generators pooling their data into a data trust.

The data trust resembles a data cooperative or a data union.[8] The trust negotiates access to the pooled data of its members, with the data service. This negotiation is part of the trust's stewardship role, and can be democratically coordinated by members. In exchange for access, the data service provides value to the trust, which is in turn passed onto members, while some is used to maintain the trust. Thus, this value is likely financial, but could also include non-financial value such as privacy rights. In the language of a union, the value used to maintain the trust would be seen not as compensation from the data service but as the data generator's contribution to the trust.

**FIGURE 3 – Data Flows in a Data Centric Data Trust[9]**




[8] The ODI (2019) distinguish between a data trust and a data cooperative by emphasising the legal structure of a data trust. Legally, such a distinction might be worthwhile. For the present purposes, however, I believe it is acceptable to consider one a variation of the other.
[9] Author's own.

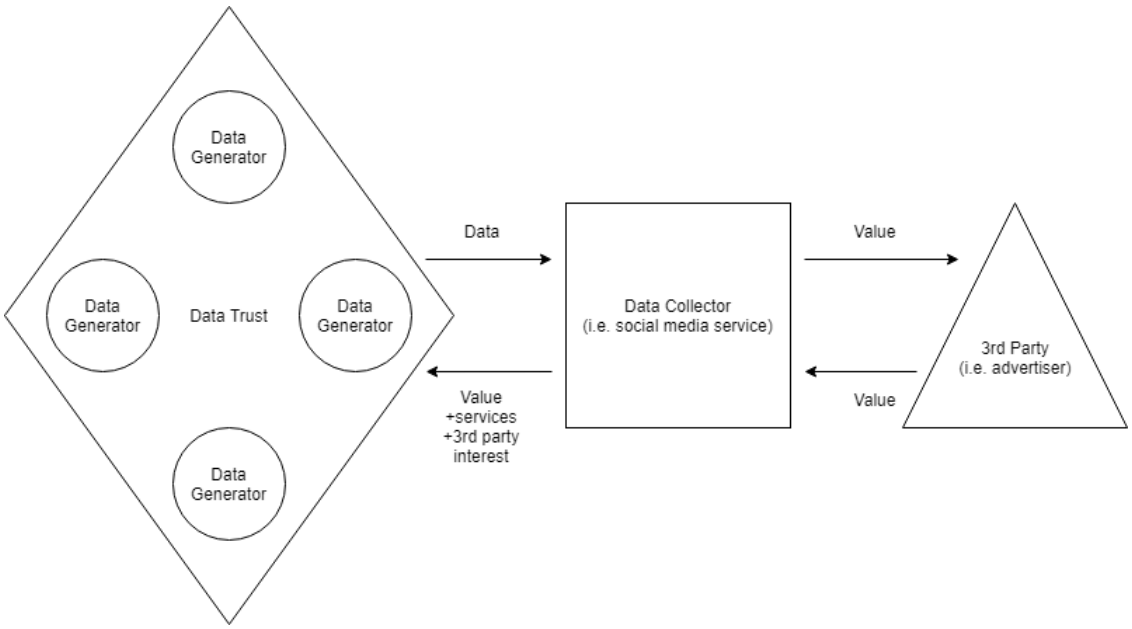
The power that is achieved by the pooling of data produces a quasi-shared decision-making relationship between the data service and generators, where the service is free to make decisions within the confines of the negotiated settlement, while the trust and its members retain the ability to withdraw access at any time (again, in the language of unions, to go on strike) if these confines are breached.

## Generator Centric Data Trust

A key problem with the data centric data trust, however, is the absence of a data collector. Returning to Gitelman (2013), data only become data when conceived as such. In a data centric data trust, it is assumed data generators collect their own data (presumably in accordance with datum requests from the data service) and the data service knows exactly what data it requires to run its service. However, it is often unclear what insights might be gleamed from data, and what uses data might ultimately come to occupy. As such, a data centric data trust – due to the lack of a collector originating new data – may find itself stuck cycling old data with innovation being restricted.

Therefore, an adjustment to the data centric data trust might be made to reintroduce a data collector – this is the generator centric data trust.

**FIGURE 4 – Generator Centric Data Trust**[10]




[10] Author's own.

The generator centric data trust resembles a much more informal union or boycott-style arrangement. This is primarily due to the appraisal of data as a resource; data are seen as an individual resource rather than pooled as in the data centric data trust. Instead of pooling data, data generator's collectivise as individuals, and the data trust governs the data collector's access to the data generator's themselves. Thus, the generator centric data trust does not steward data; instead, its stewardship obligations are the protection of individual members.

The trust is funded by value received from the data collector, which members of the trust may also receive following the same negotiated settlement seen in the data centric data trust. However, because the generator centric data trust does not control access to any data, and because it cannot compel individuals to not provide data to data collectors, the threat of withdrawing access is diminished from a union-esque strike to something more akin to a protest or a solidarity pact. A proto example of this concept can be seen in the #DeleteFacebook campaign following the Cambridge Analytica data scandal (Fowler, 2018).

However, such campaigns – and the diminished power of the trust – gleam little decision-making power away from data collectors. Instead, much like the laissez faire model, the decision-making power of data collectors is moderated only by regulation, fiduciary responsibility and reputational risk. The generator centric data trust adds to this final moderator in three ways: making the transactional nature of data production more salient; providing resources to better inform members; and serving as a centralised platform to organise members.

## Data Commons

Much like a data trust, an exact definition of a data common, as well as the features contained within that term, has yet to be determined. Brennan (2018) argues the term, 'data common,' can be viewed from a technical or governance perspective. Technically, a data common is a platform that expands data access by bringing together data from different sources.[11] This technical definition broadly fits with that given by Grossman (2018) and Grossman et al. (2016), though they also introduce an interoperability feature – a data commons is not just about access,

[11] Though this is very similar to Srnicek's (2016) definition of a platform, which is never characterised as a data common.

it's a platform for data experimentation and interaction. Grossman (2018) cites data commons in the world of healthcare as examples of this interoperability.

However, these features, nor some of Brennan's governance proposals, especially distinguish a data common from a data trust. For example, Brennan argues, "A data commons should manage permissions and control access so that it maintains the access rules and conditions under which the data were originally collected," which is very much reminiscent of the stewardship concept championed by the ODI (2019) for data trusts.

Though perhaps this quotation of Brennan (2018) is unkind, given they subsequently argue, "No data commons is an island. Each data commons should be designed to support discovering and linking to other relevant data sets, whether those data sets are held internally or located in another data commons." This feature, which Grossman (2018) supports from a technical perspective, begins to distinguish a data common from a data trust, and moves closer to what Hafen (2017) calls the, "right to a copy," (Hafen, 2017: 4) and Mills (2019), "collective transparency" (Mills, 2019: 27).

Hafen (2017) proposes an extension of human rights to include a right to a copy: the right of any person to receive a copy of any data collected about themselves. This, initially, would seem to characterise data as an individual resource, doing little to address the protestations of Yeung (2017) that data are the contrary. However, unlike a data trust which accepts Yeung's collectivised view of data and tries to preserve privacy via access controls, Hafen takes the opposite approach. Perhaps following Baudrillard's (1981) prediction that the data driven world would be, "an era of involuntary transparency," (Baudrillard, 1981: 160), Hafen suggests the copies of data held by individuals be pooled into a democratic data common.

They are not alone in this suggestion: Shah (2018) imagines data being transferred to a national data common after a period of time; Tarnoff (2018) and Morozov (2018) argue data are a collective resource and only a common can rebalance democratic control; and Srnicek (2018) goes even further, suggesting the commons model could be extended to include ownership (or at least control) of digital platforms.

Democratic oversight in Hafen's model would broadly capture the principles of stewardship, while the common structure would – they suggest – drive the value of data to zero. Theoretically, this ambition is achieved because all data services have access to the same data

(and any new data they collect is soon added to the common), and so it is not data, but the service, which generates value for these companies. Such an ambition, Mills (2019) argues, can only be achieved by accepting the principle of collective transparency, namely, data within the common must be accessible by all stakeholders. Again, this approach stands in opposition to Yeung's (2017) argument of collective privacy and raises questions regarding humanity's future right to privacy.

It does, however, offer a firm distinction between a data common and a data trust: where a data trust pools data and sets bespoke terms of access, a data common pools data and offers relatively unrestricted access.[12] This raises interesting questions regarding the (non-technical) structure of a data common and accountability, as well as the aforementioned problem of privacy.

Structurally, once more, Brennan's (2018) position seems curious. They argue that imagining a data common as a single international repository is not desirable, though provide no firm reason why, while being more sympathetic to national or industry level commons such as those discussed by Shah (2018). Yet, it is Brennan who outlines a rather compelling argument for the structural tendency of data commons in stating, "no data commons is an island." Easy access to data from various commons would suggest quick assimilation of smaller commons into an ever expanding whole. Furthermore, from an individual perspective, this single, large common seems more convenient when accessing data and sharing our individual copies of data.

Accountability is another issue a data common would need to address. Such an issue arises when considering a data common as comparable to a traditional common, immediately raising the question of how Lloyd's (1832) tragedy of the commons might be prevented. Such a tragedy is characterised as a lack of regulation which prevents initial overuse, followed by degradation, of a common resource (Ostrom, 2009). There are several ways this might occur in a data common: a high proportion of users not sharing their data; the proliferation of data practices that lead one service/user to hold undue influence; the sharing of false data; a lack of democratic oversight/participation; and pervasive knowledge asymmetries. As will be seen in Section 3.3.2, technical adjustments to the data common may ameliorate or eliminate some of

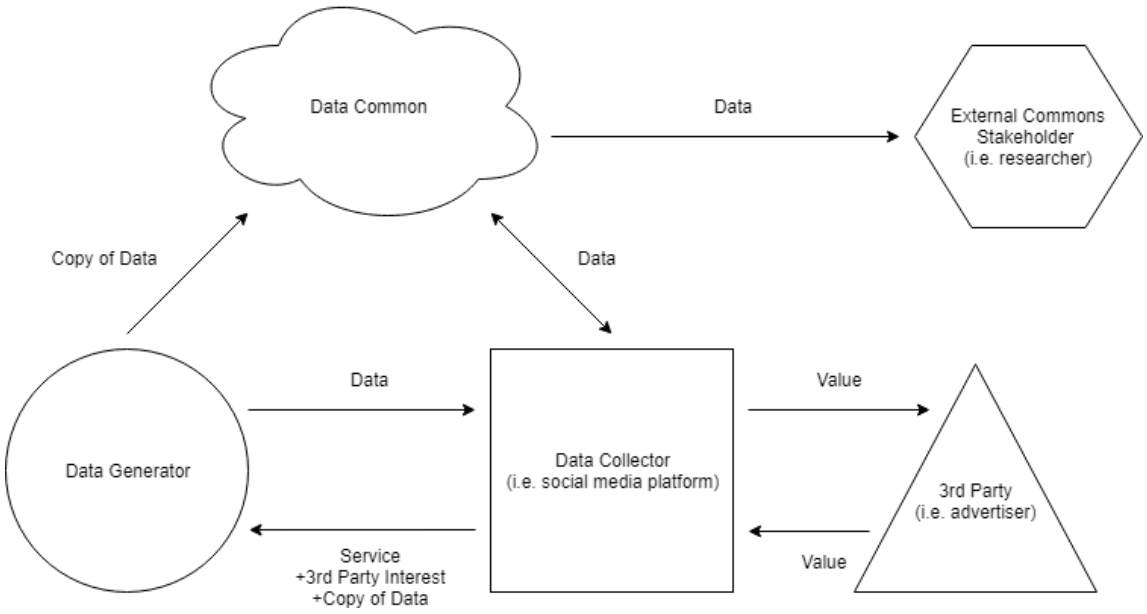
[12] If access follows the collective transparency principle, it may be permissible for a data common to grant access only after the accessor has provided their data to the common, i.e. after they have become a stakeholder in a very specific sense.

these issues. Likewise, technical adjustments may be utilised to tackle privacy concerns. Before considering these technical adjustments, however, a model of a basic data common is presented and discussed.

## Basic Data Common

Under a basic data common, the data generator and data collector interact in a similar fashion to the laissez faire model. The data collector does not necessarily have to collect data from the generator – provided it has previously been collected – because it is held in the data common which the collector has access to. However, because the maximal value of the data come when used in conjunction with the individual about whom the data apply, there is still a quasi-transactional relationship.

**FIGURE 5 – Data Flow in Basic Data Common**[13]




Data generators may choose to engage with data collectors because they provide a quality service, while collector's generate value by facilitating the access of third-party interests to generators. As such, the model fails to tackle the question of how the common is funded but does accept that value need not flow to any data generator.

[13] Author's own.

Finally, the data generator is shown in the model as a separate entity from the external commons stakeholder (who is only external from the specific data flow between generator, collector and common); however, in practice these entities can be one in the same.

## Centralised and Decentralised Data Commons

The issues outlined above, namely accountability and privacy, have been tackled in various guises by authors working on the data commons question. Lawrence and Laybourn-Langton (2018) discuss the use of national data banks and data bank accounts. Under this scheme, private-sector data would be collated into a national data common which every citizen would have an account in. The common would be run by a body designed solely for the maintenance of the common, but each citizen's data would be a matter of personal privacy, unless they chose to give it to a third party. From a technical standpoint, Grossman (2018) develops a similar concept using a globally unique ID (GUID). Under this proposal, everyone would have a unique ID which data services would have to provide in order to access data.

These proposals safeguard privacy while not undermining the purpose of the data common (to enable universal access to data). To an extent, they also temper accountability concerns, with data services having to convince individuals to provide them access to reidentified data. The centralised nature of this conceptualised data commons could also facilitate transparency (Lawrence and Laybourn-Langton, 2018), for example by mandating the governing authority publish annual statistics on which users access the most reidentified data.

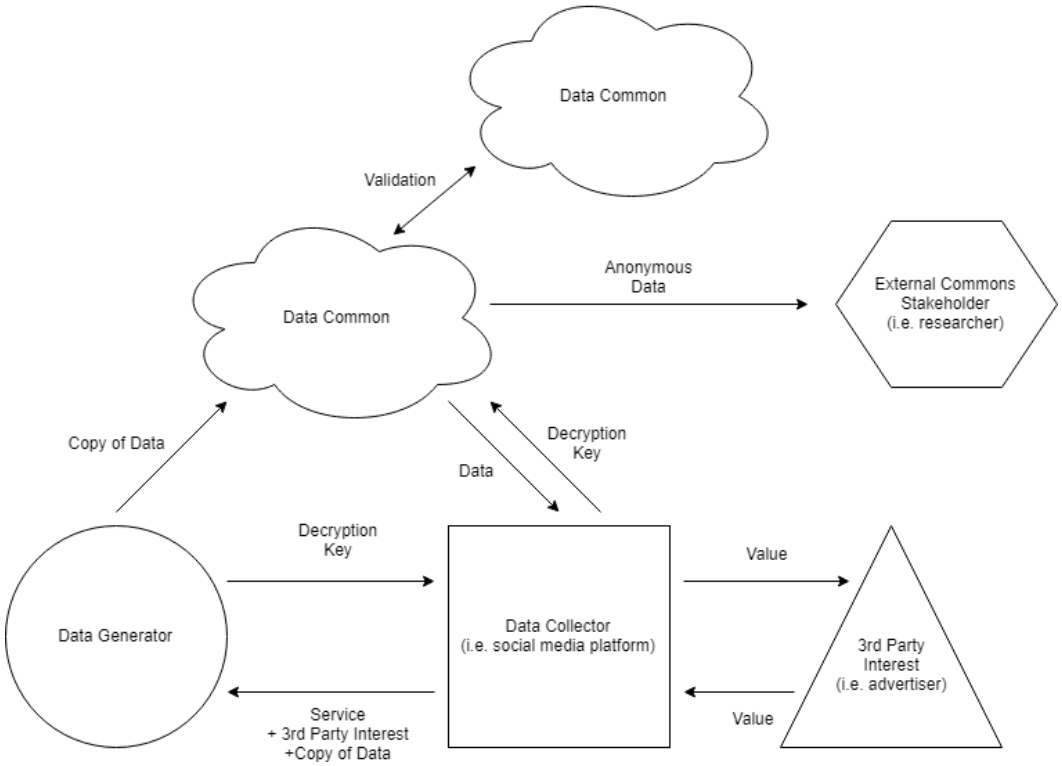
However, centralisation might also exacerbate privacy concerns, with a single breach potentially compromising all users. Furthermore, once a data service has access to reidentified data, some privacy protocols may be undermined. Finally, these systems say little about data validation, and so do not address the risk that the common becomes flooded with false or inaccurate data.

Lundy-Bryan (2018) argues a decentralised data common built on blockchain technology may address these concerns. They argue decentralisation means multiple commons exist simultaneously, allowing data on one common to be validated through comparison with several others. Data generators would add their data to a data common using an encrypted account, protecting their anonymity, while all uploads and downloads of data would be recorded on a publicly accessible blockchain, establishing accountability. Data would only become

reidentified following permission from the data generator and could be moderated using smart contracts to time limit access (Daniel and Guida, 2019). Finally, the decentralised system, Lundy-Bryan (2018) suggests, ensures no single entity controls a vast amount of data, reducing the risk of data breaches and monopoly formation.

The premise of both a centralised and decentralised data common is captured in Figure 6, which shows a data common utilising decryption keys and a validation procedure. Here, data are treated as an individual resource, but one that is pooled to prevent the formation of data monopolies that disenfranchise individual data generators. Once more, the data collector receives all the value from the data, but as with the basic data common, this value comes from the revenue they can attract due to the service they provide. However, outside of conceptualisation, it is feasible to imagine the use of encryption keys adding friction to the accessing of data, in turn allowing data to retain inherent value. However, in conceptual form, Lundy-Bryan (2018) agrees with the premise of Hafen (2017) that in a data common, the value of data falls to zero.

**FIGURE 6 – Data Flow in a Validation Data Commons**[14]




[14] Author's own.

The funding of a centralised data commons remains unclear, though reviewing the proposals of Lawrence and Laybourn-Langton (2018), it is feasible this is paid for by a national or international authority. Alternatively, data services may be compelled to pay a small fee for access, in a similar manner to how some regulatory authorities are funded in the financial industry (Financial Conduct Authority, 2019). For the decentralised data common, this approach is not possible. However, it is likely data services would assume the role in validating data held within the data common as it is in their interests to ensure they gain access to valid data. As such – again, conceptually speaking – the issue of funding a decentralised system may be reduced.

## Summary

Table 1 summarises the various models of data ownership discussed in Section 3, splitting the discussion into the three parts of the ODI's framework (ODI 2019) discussed in Section 2.

## TABLE 1 – Summary of Data Ownership Models

| Model | Creation | Stewardship | Decision-Making | Who Receives Value? |
|---|---|---|---|---|
| **Laissez Faire** | The data collector is viewed as the creator of data, and thus the de facto owner. | Stewardship is relaxed, restricted only by data protection laws, the free market and fiduciary responsibility. | The data collector has most of the decision-making authority. | Value primarily flows to data collectors and third parties. Generators receive services for free, but also third-party interests such as advertising. |
| **Data Trust:** | | | | |
| Collector Centric | Data collectors are viewed as the point of data creation. Ownership lies with the trust, though individual members may claim ownership of part of the data. | Stewardship is designed to confer monopolistic advantages to members. Third-party access has a high premium, and regulatory responsibility is shifted from collectors onto the trust. | Data collectors are bound by the rules of membership but are otherwise free to do what they desire with the data. | The administration of the trust is funded by a share of the value generated from the data. All other value flows to data collectors, while generators continue to receive value in the form of services. |
| Data Centric | Both collectors and generators are viewed as creating data. Data are seen as always being owned by generators but negotiable with collectors. | The trust stewards data on behalf of data generators. Part of this involves negotiating with collectors to ensure fair value is returned to generators. This value may be financial or non-financial, depending on the structure of the trust. | Decision-making is quasi-shared between data generators and those who access data through negotiations conducted by the trust. | Value is shared by generators and data services, depending on the settlement reached by the trust. The trust is financed by contributions from this settlement. |
| Generator Centric | Both data collectors and generators are viewed as creating data. However, ownership is generally given to data collectors, with generators coordinating to limit the power of ownership. | The trust does not control or steward data and protects members rather than their data. The trust does monitor the use of members' data by data services, may educate members, and may coordinate them. | Decision-making lies entirely with data collectors, with data seen as part of an individual transaction. However, the trust makes this transaction more salient, and so tempers decision-making with threats of boycotts and reputational damage. | While the trust may negotiate a settlement, it resembles more so a solidarity pact. As such, the cost of the trust is low. The threats levered by the trust claim some value for data generators. However, data collectors and third-parties are the primary recipients of value. |
| **Data Commons:** | | | | |
| Basic | Generators and collectors are viewed as data creators. However, data are also considered a common resource and so claims to ownership remain undefined. | Common access to data is the central stewarding principle. Degradation of the commons through the restriction of access or malpractice is also discouraged, though means of protecting the commons may not exist. | Decision-making is left to whomever accesses data. The choice to upload data is the generator's. The common is ideally structured in a democratic manner reflecting the interests of stakeholders. | Common access to data means data are not seen as a valuable resource. Value comes from services that utilise common data, and so this value is viewed as the legitimate claim of those who create it. Funding for the commons depends on structure. |
| Centralised | As above | As above. | As above, however the centralised authority may operate under different rules, and encryption techniques may give more power to individuals. | As above, though data services may pay a small fee to access data so as to fund the common, or funding may be provided by a national or international body. |
| Decentralised | As above | As above, but malpractice is prevented using a validation system, while decentralisation prevents restrictions being placed on access. | As above, however individuals have even more control over the decryption and access of their data. | As above, though the commons is supported by data services who are incentivised to perform the validation process. |

# Comparison

Throughout Section 3, clear differences both structurally and ideologically can be seen between the various models considered. In this Section 4, these differences are briefly contrasted following a (non-exhaustive) set of criteria. These are: a) incentives, b) attitude to ownership, c) competition and innovation, and d) feasibility of implementation.

## Incentives

Under the laissez faire model, collectors are incentivised to collect data via the provision of services to establish privately-held data monopolies from which they can extract the lion's share of value. This, in turn, consolidates decision-making authority in the hands of the collector.

As such, alternative models immediately disincentivise data collectors by providing data generators with a larger share of value. For example, a data centric trust – via negotiation – gleams greater value and decision-making authority from data collector's by threatening to deny them access to data. This comes with a cost for data generators, however, reducing their freedom to act as an individual agent in the market.

Not all alternative models generate disincentives for collectors, however. The exception is the collector centric data trust, which reduces the cost of collectors amassing larger datasets without compromising their, 'settlement,' with data generators. Furthermore, such coordination can produce larger data monopoly effects, entrenching incumbents and creating high barriers of entry to new entrants. Finally, this trust shifts the regulatory burden away from collectors and onto trustees, reducing administrative costs and reputational risks.

This regulatory shift is also achieved using a data common, but at a significantly higher cost for data collectors. While data commons do not necessarily demand a new value settlement between generators and collectors, the increased competition a commons creates and the diminished control data collectors experience, both represent large disincentives for collectors. Furthermore, where a great deal of a data service's value comes from its data monopoly (for example, Facebook), a data common risks massively reducing the shareholder value of these

companies. For investors, therefore, there is little incentive to support a data common, decentralised or not. However, effective coordination between data generators might eliminate the need for such approval. Generator's could then enjoy the benefits of greater market competition, though potentially at the cost of some privacy.

## Attitudes to Ownership

A loss of corporate value is just as much a reflection of attitudes to ownership as it is a disincentive. All three models of ownership have variants that continue O'Hara's (2019) neoliberal or Yeung's (2017) individualistic view of data. As data science and datasets grow, this view may become less sustainable. For example, Yeung argues larger datasets can be used to compromise the privacy of non-consenting individuals, raising both legal and moral concerns.

The collectivised view of data embraced by some data trust and data common models may rectify some of these concerns, though raise others. For example, the principle of collective transparency (Mills, 2019) and a centralised data common (Lawrence and Laybourn-Langton, 2018) may undermine individual rights to privacy and Yeung's concept of our collective right to privacy. However, a data trust with robust stewardship, or a data common with an encryption approach, may alay these fears. These fears, however, are only those of data generators. For data collectors, the collectivised view struggles to appreciate their claims to data ownership.

Finally, it should be acknowledged that much of the discussion thus far has implicitly focused on personal data. When considering natural data, such as global temperatures, a similar tension over ownership exists. However, this tension replaces data generators with the moral principle that nature belongs to everyone. As such, arguments over ownership may shift based on the type of data being claimed. It seems sensible, for example, to treat data which arise from nature as collective data, encouraging research and curiosity. However, this maxim becomes problematic when the term, 'natural,' must be defined.

## Competition and Innovation

As with ownership, competition and innovation are deeply related to incentives. It is obvious that for entrenched data collectors, competition is undesirable. Beyond a loss of value, then, a data common unleashes the competitive power of the market by greatly lowering the barriers

of entry for new participants. This, in turn, may spur innovation, with data services having to compete on the quality of their service, rather than extracting value through data monopolies. Equally, an argument could be made that – because all new data soon become available for mass use – there is little incentive for services to innovate, instead waiting for others to do so.

Thus, a data common faces a challenge in regard to innovation. Strategies such as copyrighting data or data processes may be considered, but this undermines the common, returning to a laissez faire model.

Indeed, it may be argued that the laissez faire model facilitates competition and innovation precisely because data are treated as an individual resource. This means that within the laissez faire model, effective data monopolies don't exist – rather, it is only the cost of acquiring data, not the restrictions on data themselves, which form the monopoly. This, essentially, is Srnicek's (2016) exclusion and fragmentation argument.

In contrast, while some argue data trusts also facilitate competition (Hardingess et al., 2019; Hall and Pesenti, 2017; LeCun, 2016), this article argues that competition may only be achieved if the correct data trust structure is implemented. A collector centric data trust, for example, may serve to entrench existing market players. Equally, a data centric trust may deny new entrants access to easy data, causing market stagnation. As such, a regulatory challenge like that seen with data commons emerges, with the use of government oversight or public-private partnerships becoming necessary (Young et al., 2019).

Though, it would be remiss to discuss competition without considering government. As Hall and Pesenti (2017) and Lawrence and Laybourn-Langton (2018) argue, competitive data services may prioritise easy wins over valuable data innovations. For example, transport data may be used to power a ride-hailing service but may also be used to redesign a city's transport network. The ride hailing service is incentivised to retain its data monopoly to empower its product, but there is relatively little incentive for them to use their data to aid city planners. As such, it may not be desirable for some data to be held privately in a competitive market, offering – from a regulatory perspective – an incentive to publicly claim data. Equally, access to data for public policy purposes may form the basis of public-private data sharing initiatives.

## Feasibility of Implementation

On the question of feasibility, little need be said of the laissez faire model as it is the pervasive model currently used. For data trusts, the ODI (2019) present evidence of three pilot studies, while Hall and Pesenti (2017) give an extensive, globally focused review of similar schemes. Both conclude that further development and experimentation is needed to establish a common structure and possible legal framework to empower trusts.

The exception, as proposed here, may be generator centric data trusts, with campaigns such as #DeleteFacebook demonstrating an emerging popular sentiment surrounding current data ownership issues. However, questions should be asked as to why (or whether) that movement didn't achieve its goals. One possible explanation is it lacked a formal organisational or legal structure.

Implementing data trusts, then, likely requires further investment or incentivisation to encourage experimentation. As advocated by Lawrence and Laybourn-Langton (2018), this may be backed by national executives, as private endeavours are likely to only find success implementing collector centric data trusts, as the incentives for data collectors are greatest in this model.

For data commons, the greatest implementation challenges are technical and attitudinal. A central pillar of the data common model – Hafen's right to a copy – is already enshrined in article 15 of the GDPR (O'Hara, 2019). As such, some domains will not have to face this regulatory battle.

Technically, then, the major challenge is building a platform on which the data common operates. For a centralised common, this may simply require the national executive creating a regulatory body to manage it. However, as discussed in Section 3, a centralised body creates problems – largely attitudinal – which a decentralised system may better accommodate.

These attitudinal problems revolve around privacy and authority. Many people value their personal privacy, and a major barrier to a data common will be a risk of compromise. The decentralised common (though a centralised common could also do this) uses encryption protocols to protect data, while the decentralised nature minimises the risk of these data being compromised, or invalid data being uploaded (which a centralised common may struggle with).

Of course, a decentralised common built on a technology such as blockchain faces huge technical challenges. For example, who would coordinate the building of a decentralised system, and how would the principles of governance be decided? Furthermore, many people will not understand this technology, hindering the common's adoption. However, this may be an issue experienced by all alternative models of data ownership.

## Conclusion

The Cambridge Analytica scandal is possibly the first of many to emerge as the use of data in our daily lives grows. Further, the consolidation of data in private hands possibly disadvantages regulators (Yeung, 2017) and public policy makers (Hall and Pesenti, 2017). Finally, a tension which exists between data generators and data collectors over who actually owns data is empowering movements such as #DeleteFacebook and raising questions of data's influence in civil society (Zittrain, 2013).

In response to these observations, alternative models of data ownership have begun to emerge. This article has examined three: laissez faire, data trusts and data commons, through a data value framework proposed by the ODI (2019).

I have deconstructed each model, revealing the underlying data flows and relations between actors contained within each. I have then contrasted these models by considering themes that emerge from the deconstruction: what are the incentive structures within each model; what attitude to ownership is adopted; how does the model create competition and innovation; and how feasible is the model's implementation?

It is not the intention of this article to champion any given model over another. Instead, this article's purpose is to elucidate the nuances at the heart of each model from a political economy perspective and reveal future avenues of research. As with contemporary literature such as the ODI (2019), O'Hara (2019) and Hall and Pesenti (2017), this analysis is limited in its conceptual nature. However, as further real-world experimentation is done, the challenges facing the data ownership question, and the regulation of the data industry more widely, will surely become more obvious and understood.

## Acknowledgements

I am grateful for the kind thoughts and resources provided by Jack Hardinges of the ODI, the excellent support of Daniel Bailey and Craig Berry of the Future Economies Research Centre, and the insightful feedback of Kieron O'Hara. Any errors are entirely my own.

## References

Acquisti, A, Brandimarte, L, Loewenstein, G (2015) 'Privacy and Human Behavior in the Age of Information' *Science*, 347(6221), pp. 509-14

Beer, D (2017) 'The Social Power of Algorithms' *Information, Communication and Society*, 20(1), pp. 1-13

Bastani, A (2019) '*Fully Automated Luxury Communism: A Manifesto'* Verso Books: London

Baudrillard, J (1981) '*Simulacra and Simulation'* Glaser, S (1994) Michigan University Press: USA

Brennan, P (2018) '*What makes a data commons work?*' NLM Director. [Online] [Date accessed: 14/08/2019]: https://nlmdirector.nlm.nih.gov/2018/04/24/what-makes-a-data-commons-work/?platform=hootsuite

British Academy and Royal Society (2017) '*Data Management and Use: Governance in the 21st Century*' [Online] [Date accessed: 13/08/2019]: https://www.royalsociety.org/topics-policy/projects/data-governance

Confessore, N (2018) '*Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*' The New York Times. [Online] [Date accessed: 11/08/2019]: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Coyle, D, Nguyen, D (2019) 'Cloud Computing, Cross-Border Data Flows and New Challenges for Measurement in Economics' *National Institute Economic Review*, 249(1), pp. 30-38

Daniel, F, Guida, L (2019) 'A Service-Oriented Perspective on Blockchain Smart Contracts' *IEEE Internet Computing*, 23(1), pp. 46-53

Determann, L (2018) '*No One Owns Data'* SSRN. [Online] [Date accessed: 26/08/2019]: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3123957

Economist (2017) '*The world's most valuable resource is no longer oil, but data'* The Economist. [Online] [Date accessed: 11/08/2019]: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

European Parliament (2016) '*Regulation (EU) 2016/679 of the European Parliament and of the Council'* Official Journal of the European Union. [Online] [Date accessed: 13/08/2019]: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Financial Conduct Authority (2019) '*About the FCA' Financial Conduct Authority'* [Online] [Date accessed: 14/08/2019]: https://www.fca.org.uk/about/the-fca

Forbes (2019) '*The World's Most Valuable Brands'* Forbes. [Online] [Date accessed: 11/08/2019]: https://www.forbes.com/powerful-brands/list/#tab:rank

Fowler, G (2018) '*Go ahead and #DeleteFacebook. But here's the change we really need'* The Washington Post. [Online] [Date accessed: 13/08/2019]: https://www.washingtonpost.com/news/the-switch/wp/2018/03/21/go-ahead-and-deletefacebook-but-heres-the-change-we-really-need/

Furman, J (2019) '*Unlocking Digital Competition: Report of the Digital Competition Expert Panel'* [Online] [Date accessed: 13/08/2019]: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf

Gitelman, L (2013) '*Raw Data Is an Oxymoron'* MIT Press: Boston

Grossman, R (2018) '*A Proposed End-To-End Principle for Data Commons'* Medium. [Online] [Date accessed: 14/08/2019]: https://medium.com/@rgrossman1/a-proposed-end-to-end-principle-for-data-commons-5872f2fa8a47

Grossman, R, Heath, A, Murphy, M, Patterson, M, Wells, W (2016) 'A Case for Data Commons: Toward Data Science as a Service' *Computing in Science and Engineering*, 18(5), pp. 10-20

Hall, W, Pesenti, J (2017) '*Growing the Artificial Intelligence Industry in the UK'* UK Government. [Online] [Date accessed: 13/08/2019]: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

Hardinges, J, Wells, P, Blandford, A, Tennison, J, Scott, A (2019) '*Data trusts: lessons from three pilots'* Open Data Institute. [Online] [Date accessed: 11/08/2019]: https://theodi.org/article/odi-data-trusts-report

Helbing, D, Frey, B, Hafen, E, van den Hoven, J, Gigerenzer, G, Zicari, R, Zwitter, A,

Hofstetter, Y (2017) '*Will Democracy Survive Big Data and Artificial Intelligence*?' [Online] [Date accessed: 10/07/2019]: https://www.researchgate.net/publication/325105787_Will_Democracy_Survive_Big_Data_and_Artificial_Intelligence

Hodgskin, T (1825) '*Labour Defended Against the Claims of Capital'* (1997) Routledge: London

International Accounting Standards Board (2018) '*Conceptual Framework for Financial Reporting'* [Online] [Date accessed: 13/08/2019]: https://www.accaglobal.com/content/dam/acca/global/pdf/sa_mar11_f7p2.pdf

Isaac, M (2019) '*Mark Zuckerberg's Call to Regulate Facebook, Explained'* The New York Time. [Online] [Date accessed: 11/08/2019]: https://www.nytimes.com/2019/03/30/technology/mark-zuckerberg-facebook-regulation-explained.html

Kitchin, R (2014) '*The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*' Sage: London

Lanzing, M (2018) '"Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies' *Philosophy and Technology*, 30(2), pp. 1-20

Lawrence, M, Laybourn-Langton, L (2018) '*The Digital Commonwealth: From private enclosure to collective benefit'* IPPR Commission on Economic Justice. [Online] [Date accessed: 14/08/2019]: https://www.ippr.org/files/2018-08/cej-platforms-sept18.pdf

LeCun, Y (2016) '*Artificial Intelligence in the Real-World'* Economist Intelligence Unit Briefing Paper (2016). [Online] [Date accessed: 13/08/2019]: https://eiuperspectives.economist.com/sites/default/files/Artificial_intelligence_in_the_real_world_1.pdf

Lloyd, W F (1832) '*Two Lectures on the Checks to Population'* University of Oxford: Oxford. [Online] [Date accessed: 14/08/2019]: https://philosophy.lander.edu/intro/articles/lloyd_commons.pdf

Lundy-Bryan, L (2018) '*A Real Use Case for Blockchains: A Global Data Commons'* Outlier Ventures. [Online] [Date accessed: 14/08/2019]: https://outlierventures.io/research/a-real-use-case-for-blockchains-a-global-data-commons/

Marr, B (2018) '*Here's why data is not the new oil'* Forbes. [Online] [Date accessed: 26/08/2019]: https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/

Mazzucato, M (2018) '*Let's make private data into a public good'* MIT Technology Review. [Online] [Date accessed: 11/08/2019]: https://www.technologyreview.com/s/611489/lets-make-private-data-into-a-public-good/

Mills, S (2019) '*Into Hyperspace: The Challenges Facing Big Data and Hypernudges'* SSRN. [Online] [Date accessed: 13/08/2019]: intodhttps://papers.ssrn.com/sol3/papers.cfm?abstract_id=3420211

Morozov, E (2018) '*After the Facebook scandal it's time to base the digital economy on public v private ownership of data'* The Guardian. [Online] [Date accessed: 14/08/2019]: https://www.theguardian.com/technology/2018/mar/31/big-data-lie-exposed-simply-blaming-facebook-wont-fix-reclaim-private-information

O'Hara, K (2019) '*Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship'* WSI White Paper #1. [Online] [Date accessed: 11/08/2019]: https://eprints.soton.ac.uk/428276/1/WSI_White_Paper_1.pdf

Ostrom, E (2009) '*Governing the Commons'* Cambridge University Press: London

Rees, C (2013) '*Who Owns Our Data*?' SSRN. [Online] [Date accessed: 26/08/2019]: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2310662

Scassa, T (2018) '*Data Ownership'* CIGI Papers No. 187, SSRN. [Online] [Date accessed: 26/08/2019]: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3251542

Schlosser, A (2018) '*You may have heard data is the new oil. It's not*.' World Economic Forum. [Online] [Date accessed: 26/08/2019]: https://www.weforum.org/agenda/2018/01/data-is-not-the-new-oil/

Shah, H (2018) 'Use our personal data for the common good' *Nature*, 556(7), pp. 7-8

Srnicek, N (2016) '*Platform Capitalism'* Polity: UK

Srnicek, N (2018) 'Technology, Capitalism and the Future of the Left' *Renewal: A Journal of Social Democracy*, 26(1), pp. 18-31

Tarnoff, B (2018) '*The Data Is Ours*!' Logic. [Online] [Date accessed: 14/08/2019]: https://logicmag.io/scale/the-data-is-ours/

United Nations (1948) '*Universal Declaration of Human Rights'* [Online] [Date accessed: 13/08/2019]: https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf

Van Dijck, J (2014) 'Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology' *Surveillance and Society*, 12(2), pp. 199-208

Warren, E (2019) '*Breaking Up Big Tech'* [Online] [Date accessed: 24/09/2019]: https://elizabethwarren.com/m-break-up-big-tech/

Yeung, K (2017) 'Algorithmic Regulation: A Critical Interrogation' *Regulation and Governance*, 12(4), pp. 505-523

Young, M, Rodriguez, L, Keller, E, Sun, F, Sa, B, Whittington, J, Howe, B (2019) '*Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing'* FAT'19 Conference Paper. [Online] [Date accessed: 14/08/2019]: https://faculty.washington.edu/billhowe/publications/pdfs/young_open_v_closed_semi_synthetic_data.pdf

Zittrain, J (2013) '*Engineering an Election'* [Online] [Date accessed: 11/08/2019]: http://cdn.harvardlawreview.org/wp-content/uploads/2014/06/vol127_Symposium_Zittrain.pdf

**Future Economies**
**Business School**
**Manchester Metropolitan University**
**Oxford Road**
**Manchester**
**M15 6BH**

**https://www2.mmu.ac.uk/future-economies**

**https://twitter.com/mmu_futureecon**