SCIENCE POLICY RESEARCH UNIT

# SPRU Working Paper Series

SWPS 2019-21 (October)

Research Note

## The Value of Data: Towards a Framework to Redistribute It

Maria Savona

Energy Policy

Sustainable Development

Science and Technology Policy

Innovation and Project Management

Economics of Innovation

UNIVERSITY OF SUSSEX | BUSINESS SCHOOL

# SPRU Working Paper Series (ISSN 2057-6668)

The SPRU Working Paper Series aims to accelerate the public availability of the research undertaken by SPRU-associated people, and other research that is of considerable interest within SPRU, providing access to early copies of SPRU research.




# Guidelines for authors

Papers should be submitted to swps@sussex.ac.uk as a PDF or Word file. The first page should include: title, abstract, keywords, and authors' names and affiliations. The paper will be considered for publication by an Associate Editor, who may ask two referees to provide a light review. We aim to send referee reports within three weeks from submission. Authors may be requested to submit a revised version of the paper with a reply to the referees' comments to swps@sussex.ac.uk. The Editors make the final decision on the inclusion of the paper in the series. When submitting, the authors should indicate if the paper has already undergone peer-review (in other series, journals, or books), in which case the Editors may decide to skip the review process. Once the paper is included in the SWPS, the authors maintain the copyright.


Websites

UoS: **www.sussex.ac.uk/spru/research/swps**
SSRN: **www.ssrn.com/link/SPRU-RES.html**
IDEAS: **ideas.repec.org/s/sru/ssewps.html**

# The Value of Data: Towards a Framework to Redistribute It

[Professor Maria Savona](#)

Science Policy Research Unit, University of Sussex

M.Savona@sussex.ac.uk; @maria_savona

First version: July 2019; Second version: October 2019. Preliminary, comments welcome

## Abstract[1]

This note attempts a systematisation of different pieces of literature that underpin the recent policy and academic debate on the value of data. It mainly poses foundational questions around the definition, economic nature and measurement of data value, and discusses the opportunity to redistribute it. It then articulates a framework to compare ways of implementing redistribution, distinguishing between data as capital, data as labour or data as an intellectual property. Each of these raises challenges, revolving around the notions of data property and data rights, that are also briefly discussed. The note concludes by indicating areas for policy considerations and a research agenda to shape the future structure of data governance more at large.

[1] I wish to thank first and foremost Andrea Glorioso for the excellent discussions on the topic and comments on previous versions of this note. Mr Glorioso wishes to note that his comments did not necessarily reflect the official position of the European Union or of any of its Institutions and Bodies. I warmly acknowledge suggestions by the SWPS reviewers. I am also very grateful to Ed Steinmueller, Simone Vannuccini and Ariel Wirkierman, for remarks and suggestions on a previous draft; to Alberto Marzucchi and Tommaso Ciarli, within the joint organisation of the EMAEE 2019 Conference on the Economics, Governance and Management of AI, Robots and Digital Transformations held at SPRU in June 2019. The conference debate between Carol Corrado and Jennifer Morone on "Data as Capital or Data as Labour?" has been an invaluable source of reflections. I am grateful to the participants to the Knowledge Symposium, University of Heidelberg, 27 June 2019 where I presented "Digital Transformation and Digital Ownership. Policy Challenges to redistribute data value", particularly to Michael Handke; and the participants to the Alan Turing Workshop on Data Trusts, Cambridge Jesus College, 23 September 2019, for the great debate on related and relevant issues. I had fruitful exchanges with Luca De Biase at the EY Digital Summit held in Capri in October 2019, where we debated around the proposal and challenges illustrated here. I am greatly indebted to Katherine Davies, who has provided excellent editing and suggestions to my blog and written pieces around the theme. All omissions are my responsibility.

## 1. Introduction

Digital transformations are creating great opportunities but also challenges for modern labour markets. Supporting and steering transformations while reducing disruption is a pertinent policy challenge. While discussions must address the technological anxiety and declining working conditions associated with advances in AI and automation, they should also consider how digital transformations could reduce unemployment and underemployment, and increase prosperity and inclusion at the European level.

To explore possible solutions to these policy challenges, the European Commission in September 2018 convened an expert group on the impact of the digital transformation on EU labour markets. The group included experts from the public and private sector, alongside a few academics. The HLG convened over five monthly meetings that provided opportunities to cross-fertilise our own expertise. We were asked to think outside the box, providing ground-breaking policy recommendations based on empirical evidence. These have fed into a final report, released on 8 April 2019 and discussed in Brussels at the High-Level Conference on the Future of Work. The nine recommendations are summarized on the expert group's home page.

This document develops the background underpinning the recommendations on redistribution of data value that I put forward in the report, with the aim of systematising the positions emerged in less and more recent literature, articulating them in further detail, proposing alternative ways of implementation and discussing the challenges arising. Up to Section 3, I abstain from considering whether the concept of data ownership is appropriate or not in this context.

The economic value of data[2] has increased substantially in recent years. Advances in automation and Machine Learning (ML) rely heavily on feeds of data to transform into relevant intelligence. Currently, the necessary infrastructure to treat, process and analyse these data (and thus benefit from their value) is mainly concentrated in a handful of big tech companies. However, data-value is largely dependent on the constant feed of *personal data*

[2] Unless otherwise specified, here I refer to data mainly as personal data shared by individuals in the context of their leisure or work activities online.

generated by (online) consumers and workers within companies and on the possibility to store, aggregate and treat individual data.

The European Commission has been at the forefront to promote convergence in the global governance of data (privacy), including, but not limited to, the well-known EU GDPR (General Data Protection Regulation). Despite these efforts, the current (or currently absent) global data governance structure calls for a discussion **of how data value is (or should be) distributed**, which aims to go beyond simply revisiting traditional competition policy.

So far, the collection and appropriation of data by companies has gone largely unquestioned. People willingly provide their personal data in exchange for using an online service, and perceive this exchange as a barter. And firms also benefit from collecting their employees' and consumers' data, uncompensated. But there is an increasing need to consider – and measure – the value of data.

Attempts to do so are not new. Companies' capabilities to innovate and grow are determined not only by their investments in R&D, training, engineering, design and so on, but also their ability to aggregate, treat and learn from the data they are gathering (see OECD, 2018 and previous versions of the Oslo Manual). Advancements in data development and analytics mean that such information are increasingly seen and measured as 'knowledge-based capital' or 'intangible assets.' Measurement and economic impact of intangibles have been the object of an established branch of literature (see Corrado et al., 2009 among others), although not specifically to the emergence of big data.

Policies to redistribute data value would need a radical rethinking of the nature of data in the current technological landscape, and within the extant narrative around intangible investments and intangible capital.

## 2. Foundational questions

The concept of bartering personal data for online services counterposes a view in which individual data contributions can be seen as the initial stage of a **data value chain** and, as such, worthy of remuneration.

On the one hand, data generated, aggregated and treated within firms has led to standard practices of workers not being directly compensated for sharing their private data with firms. On the other hand, data aggregation, treatment, development and analytics – alongside data management skills – are included among what economists consider as **intangible assets** of firms. These traditionally contribute to the knowledge-based capital in national accounts (alongside Research & Development (R&D), Intellectual Property Rights (IPR), training, software, engineering and design, marketing and branding), as their collection, stocking and analytical treatment entail investments from firms.

The practice has superseded the theory in this case, and the literature on intangible assets has contributed to making the practice an unquestioned standard. The practice of carrying out data analytics and firms' investment in transforming this into data intelligence (thus an intangible asset) has led to a high concentration of equity value in a few companies. This is based on their retaining not only direct access to the source of data, but also the infrastructures to create incentives to and influence individual data generation and the emergence of the very markets that buy data intelligence.

Thus far, the measuring of intangible capital has incorporated a broad range of items from R&D investments, to trademarks, to human capital, alongside data. This practice closely aligns with the guidelines for measurement of innovative investments provided by the Oslo Manual (OECD, 2018). There is, however, the question of whether the value generated by these different items can be considered homogenous, at least to the extent that they can be safely included in the same (intangible) *capital stock*.

Ideally, we could ask why a firm paying a licence for the use of externally-sourced knowledge, (i.e. patents or trademarks owned by other firms), or indeed receiving a licence for the use of their patents or trademarks, would include these as intangible investments, while it does not necessarily pay for the use of individual data generated by consumers or workers (similarly external sources of information). Such data can then be aggregated and treated with data analytics to generate value. Both types of investments are **intermediates** to the generation of intangible value, both are **sourced externally**, both need to be **further treated** by the typical firm to generate value.

Including different items into the stock of intangible capital is methodologically sound only if we use the same criteria (and price system to start with) to track the generation of value, from raw data as intermediates that enter the data value chain to the cumulated value of data intelligence.

The appropriation of external (intermediate) value is remunerated in the case of external R&D or trademark acquisition, but is not in the case of individual data.

*Why?*

We should step back and ask several foundational (and thus far unanswered) questions that might change how we approach data value. These pertain to different realms and arise at different stages. We first revert to the economic nature of a 'good' (Box 1), then problematise the nature of data (Box 2), and reflect on whether we need a policy to redistribute data value and what could be a framework to tackle this (Box 3).

*Economic nature of data and implications for measurement of data value*

**Box 1 – Excludability, Rivalry and types of goods**

**Excludability:** a good is excludable if an individual can be denied access to or consumption of it.

**Rivalry:** a good is rivalrous if its consumption by an individual makes it used up or unsuitable for use by someone else.

The economic nature of goods in terms of excludability and rivalry is determined both by law, which regulates the contractual rights over their ownership and use, and by the scale of their supply.

**Public Goods** are both non-excludable and non-rivalrous.

**Private Goods** are typically excludable and rivalrous.

**Club Goods** are a particular set of public goods, as they are non-rivalrous, but excludable as individuals can be denied access to or use of them.

**Common Goods** are goods that are non-excludable and rivalrous, typically natural resources, whose use is limited (rivalrous) but are not excludable.

Based on the notions of excludability and rivalry, some questions arise on the specific nature of data, summarised in Box 2 below.

**Box 2 – Measurement of data value**

What is the economic nature of data? (*public, private, club or common good*)

Which type(s) of data are valuable?

How is data value created, and extracted?

How is data value currently tracked and measured, if at all?

Is it possible to unambiguously identify the stages of a data value chain, to recognise data value generators and data value extractors along different stages of the data value chain?

If it is, what price system is best fit to value data and quantify the concentration of data value?

It has been observed (Jones and Tonetti, 2019) that data are typically non-rival and therefore can be considered as club goods. However, because of the very nature of data, particularly the production process that entails a data value chain, their nature might not be as static.

Before bringing all the pieces together, though, a set of different questions arises, that should be grounding any reasoning around the governance of data value (Box 3).

*(Re)distribution, Governance and Challenges*

**Box 3 – Governance of data value redistribution: main challenges**

What types of data raise an issue of value redistribution? *Personal data for private use? Personal data for public use? Both, regardless of their use destination?*

Is the concept of data ownership subsumed under the process of data value appropriation?

Is there such a thing as a public value of data, and is data use by state enterprises or other organs of the state legitimated as it carries a public value?

Can we devise a framework to redistribute corporate equity value generated by accessing individual data value, beyond corporate taxes?

Can this framework bypass the controversial concept of personal data ownership?

Which institutions are best placed to implement processes of redistribution of data value?

Are there fundamental ethical issues raised by the framing of an individual as the portable bundle of data (s)he generates?

Does this framing align with the rationale underpinning data privacy protection as a fundamental right (as in GDPR)?

Data subjects create value individually. Data value is then extracted by data capitalists through investments in digital infrastructures, organisational and human capital, enabling data collection and aggregation, treatment and analysis. How should policies redistribute the data value? By compensating data value creators and/or taxing data value extractors?

In what follows, we consider some of the above questions and put forward a framework to initiate discussion around policies for data value redistribution. In a nutshell:

a) Should we consider data as **intangible capital** and simply tax more effectively the profits generated by data ownership and analytics?

b) Should we consider **data as labour** and rethink the configuration of the labour market, for instance offering a wage premium to workers who generate data appropriated by companies?

c) Should we consider data as the **Intellectual Property (IP)** or more simply a **licensable asset** owned by the individual who generates it, as part of a system that recognises and protects intellectual property rights and asks platforms owners to pay a licence for use?

I will argue that there are pros and cons with each of these, particularly when it comes to resolve data value fairness. I will attempt to propose a different thinking around data, which might bring about more benefits than a more traditional regulation of data brokers and data markets.

## 3. A framework to redistribute data value: Data as Labour, Capital, Intellectual Property

### A taxonomy of the nature of data

Data are a polymorphous category. Different types of data enter the chain of data value, and possibly only some of them are arguably valuable.

Data can be valuable *for both data generators and data brokers in similar ways*, as is implicit in the idea of a 'barter' between online platforms and their users/consumers, in which a service is delivered in exchange for individual data provided (for free).

Some data only have a *use value* while other have clearly an *exchange value.* Some data might become obsolete as they provide obsolete information, and their value could decline.

A good starting point is to devise a preliminary taxonomy to identify which data are (or should be) the object of a redistribution policy.

If we reprise Buchanan (1965) we could apply his definition of 'club goods'[3] to most categories of personal data that are appropriated by companies. Personal data generated as ***information*** (vis à vis *knowledge*, below) are *excludable but non-rivalrous* as any club good. In other words,

[3] As illustrated in Box 1, club goods are typically non-rivalrous, i.e. their consumption by an individual does not make them unsuitable for use by another individual and their size does not typically lead to their being used up; but, unlike pure public goods, they are also excludable, i.e. their use or access can be denied.

big (personal) data have high excludability but very low rivalry as their scale is immense. Because of their low rivalry in consumption, club goods have a *zero-marginal cost*. In the case of data, the high fixed costs of the digital infrastructure to collect and treat data, combined with the zero marginal costs of individual data, is what make big tech companies quasi-natural monopolies.[4] There is therefore a typical asymmetry in the value of data for members of the club, i.e. data generators and data collectors. However, **devising a price system for goods that have zero marginal costs is a difficult (and useless?) endeavour**.

Once the *individual data* are aggregated, and treated to generate value from data analytics and relevant intelligence, their nature changes from that of club goods.

Data analytics becomes ***knowledge,*** i.e. a *public good (non-rivalrous and non-excludable).* Data intelligence is a public good if it is used for non-profit purposes and managed by public actors (examples include the use of biometric data for health research). When information is processed and becomes knowledge, its nature of public good might (or might not) lead to public benefit, as long argued by economists of innovation (among others, Foray, 2004).

However, when appropriated by private actors for profitable purposes (as in marketing or other forms of profit-led intelligence) they become a *private good (excludable and rivalrous).*

It is a matter of policy choice to identify which category of data should be the object of a redistribution policy. The framework below – for simplicity – applies to any category of individual data, appropriated by public or private actors through collection, storage, aggregation and treatment for public or private purposes. I delineate three ways to frame a redistribution rationale, which could be alternative or complementary.

---

[4] Within a marginalistic approach to data value, a zero marginal cost means a marginal value of data close to zero. However, data have increasing returns to scale rather than decreasing ones. For instance, according to Posner and Weyl (2018), the machine learning complexity-based marginal value of data grows as a function of the number of tasks you want to accomplish. In the authors' own words: "(..) the primary determinant of the marginal value (of data) is not the statistics of a given ML problem, but rather the *distribution of the complexity across different problems"* (Posner and Weyl, 2018, p. 228).

## Data as Capital

One way is to consider **Data as Capital (DaC).** Prior to the rise of 'Big Tech' in the last decade, companies relied on intangible capital and innovative capabilities for a relatively longer period of time. Investments in R&D, patents, trademarks, software, design, engineering, training, and organisational capital all fed into their 'knowledge-based capital' or 'intangible assets' and were measured as such (Corrado et al., 2009). Data collection, treatment and analytics represent intangible investments and have only more recently started representing the lion's share of accumulated collective intelligence on which companies rely.

Interestingly, Corrado (2019) considers a **Data Value Chain** in which raw individual data bear the lowest value, while what adds value are the processes that aggregate, systematise and treat raw individual data. In this system, the main outcome of data analytics becomes intelligence that is appropriated by the company. The literature on intangibles (Corrado et al., 2009) considers the spending on data analytics, now increasingly Machine Learning (ML) and Artificial Intelligence (AI), as intangible investments and measures them as such. Within national accounts, data analytics are covered in the methods that are used to estimate software investments. These have two possible sources: the first is "own account software", the value of which is measured by the compensation of work done by employees in certain tasks, such as engineers and programmers (Brynjolfsson et al., 2018); the second is "purchased software", acquired from relevant external markets (Corrado, 2019).

A central underpinning of the literature on intangibles is based on the current US accounting system, which has long considered datasets as intangible assets, similar to the way they are considered in the OECD Oslo Manual guidelines for measuring innovation investments.[5] Certainly, some of the national accounting elements need to be updated to incorporate aspects that are unique to the current AI systems. In fact, some available indicators suggest that - while the growth rate of investments in AI is increasing in the US – the investment rate

[5] In this context, it is worth noting that owning intangibles assets is not necessarily a synonymous of being innovative. One might argue that the stock of personal data for profitable purposes such as marketing profiling is not equivalent to capital stock of R&D investments for innovation purposes. Taxing profits generated by data analytics is legitimate as long as it is not discriminatory against R&D intensive firms, which are expected to create knowledge and societal spillovers from their research investments.

in intangibles on the whole is decreasing (Corrado, 2019). Recent work (Brynjolfsson et al., 2017) raises the challenges posed by AI when measuring national intangible investments.

### Redistributing value within the "Data as Capital" framework

If we frame the redistribution around the notion of "Data as Capital", the most straightforward way to redistribute the value of the collective intelligence from mass-scale data would be to rely on supranational public institutions that create the (so far missing) market for data (Ibarra et al., 2018). One could then design an adequate system of taxation of current data owners, similar to the 'bit tax' (Soete and Kamp, 1997) (proposed over twenty years ago) or the more recent 'robot tax' (as proposed to the European Parliament by the Committee of Legal affairs[6]). European Union law on the protection of personal data, including the European General Data Protection Regulation (GDPR), can be considered a fundamental milestone in data regulation, which prepares the ground for balancing the power between data generators and data extractors. The GDPR would, in principle, be consistent with shifting the rights of data (value) creation to the individuals that generate them. This proves that concerted government action can play an important role towards building a new regulatory framework that deals with the (market) failure of a missing market for data.

The fact that data are included as intangible assets in firm balance sheets could facilitate the practical implementation of a tax on intangibles capital by fiscal authorities. This is however based on the heroic assumption that, once released by the individual, data are easily trackable and it is possible to reconstruct a data value chain, based on an appropriate price system to quantify costs of storage, aggregation and treatment. Devising corporate taxes on intangible capital is therefore not obvious. In addition, the very nature of data would make the role of supranational fiscal institutions more appropriate, in a context of increasingly undermined traditional national tax bases.

More in general, the usual resistance to capital taxes is that governments are generally unwilling to impede capital formation which is the source of growth and employment as well

[6] European Parliament - Document/A-8-2017-0005. See also for a response R. Viola, Robotics will be a key driver of economic growth, The Parliament Magazine, December 2017.

as profits. Better to tax the golden eggs than the golden goose lest the owner decide not to allow the goose to reproduce. One might argue, again, that the big techs are not exactly – or not always - the golden gooses, the same way a firm highly investing in R&D would be.

## Data as Labour

Another option is to treat **Data as Labour (DaL),** meriting (wage) remuneration. According to Ibarra et al., (2018) (see also Posner and Weyl, 2018), there are several possible advantages of treating individual data as labour, including reducing the need to redistribute income by imposing a corporate tax on digital activities, as illustrated earlier. [7]

An emergent debate, stemming from the RadicalxChange foundation in the US, reprises Posner and Weyl (2018), who argue that the “powerhouse of the digital economy” – the Big Tech – exploit the lack of public understanding of how machine learning (ML) and AI collect and treat data that are generated by consumers and workers.

The (not too) science-fiction scenario described in Posner and Weyl (2018) is that the mass unemployment feared as a result of AI is a myth, as humans are still fundamental as data generators, the work of the future. The missing labour market of data generators, Posner and Weyl (2018) argue, could be likened to the unpaid housework performed by homemakers. A ‘data dignity’ narrative is then developed, borrowing from the notion of ‘labour dignity’ (Morone, 2019).

### Redistributing value within the “Data as Labour” framework

Starting to remunerate even a small subset of high-value data for specific purposes would radically and irreversibly change big platforms’ business models (undermining their oligopsony power). The digital commons currently have very little space for competitors.

[7] Ibarra et al., 2018 argue for instance that treating data as labour could Increase the productivity of AI systems; encourage entrepreneurship and innovation by individuals, leading to an increase in the quality and quantity of data and be a source of self-esteem in a context where the changing nature of work is expected to reward the individual beyond pure financial remuneration.

The challenges of implementing this process are, however, significant. First, a substantial overhaul would be needed to adequately inform consumers – and, indeed, workers - of the unequal terms of the barter they are currently offered before feeding their data into big tech platforms. Second, and no less important, a ‘general purpose’ tracking infrastructure does not yet exist and would have to be imposed by public regulation, which raises the same issues that we have highlighted in the previous section. Also, this approach would need a radical rethinking of the labour markets configuration, including mitigating the potential for counterproductive exploitation and misuse of the incentive to generate unnecessary mass of data within labour contracts.

To address this, the development of MIDs (Mediators of Individual Data) has been proposed. MIDs work as intermediaries between individual data generators and companies, and their role is to mediate data exchange. They should be fiduciary, ensure high standards of data quality, make data provenance inalienable, allow equal, and possibly long term, sharing of benefits (Morone, 2019). MIDs act as cooperative data brokers and allow for collective or communal data ownership by the individual data generators. Despite their nature of data commons, MIDs likely need government support to take off. We will return to this later in Section 3.

As part of the ‘data dignity’ plea, Morone (2019) argues, the role of unions would represent a countervailing power to that of the big tech oligopsony. Data labour movements and digital unionisation could, in a similar way to the 20$^{th}$ century industrial union, allow: (i) Collective bargaining (data labourers are incredibly dis-homogeneous since the de-industrialisation process started to disintegrate a collective identity of (factory) workers); (ii) quality certification (data quality should be ensured and misuse of incentives mitigated perverse incentives avoided); (iii) career development. Each of these points would deserve a thorough reflection and possibly more grounding.

Interestingly from our (academic) perspective, the “Data as Labour” framework should engage with debates around the changing nature of work, particularly on what the *intrinsic* and *extrinsic* incentives to work (Bénabou and Tirole, 2006) have become in the digital economy. Remunerating data generators could either undermine or enhance the social rewards attached to the sense of belonging to online communities, of being empowered and contributing to social value, i.e. the *intrinsic* incentives. With billions of heterogenous potential digital workers, this is however hard to predict. If providing good quality data can increase workers' self-esteem, then remuneration would be based on a grounded rationale. If, instead, it increases their *extrinsic (perverse)* incentives to provide low quality data to maximise their financial rather than social rewards, then a "Data as Labour" framework would not be optimal.

Also, again from the (academic) perspective, it would be of extreme interest to re-visit the theory of value in the economic discipline, to account for digital transformation and what capital and labour have become within these. From such a perspective, the "Data as Labour" framework seems to substantially converge towards the "Data as Capital" framework as they both rely on a pricing system that risks being fundamentally detached from the actual value system. Also, if data are the new labour, can we revisit the theory of labour value? And, relatedly, are we trying to devise a method of valuing an intangible product (data) or remunerating the source of its (re)production?

### Data as an Intellectual Property and a licensable asset

A third, novel, and possibly more inclusive way to tackle the issue of redistributing data value is to treat workers' and consumers' data **as Intellectual Property[8], worthy of remuneration as a usable and licensable asset and of protection as an Intellectual Property Right ("Data as Intellectual Property Right")**. Data generated by both workers (within a labour contract, and through the process of carrying out their tasks) and consumers (outside the firm, but through the process of consuming services and thus appropriated by the firm) are owned by them, **to the extent that they result from their individual investments in their own personal knowledge**. Data can therefore can be treated as intellectual property and be protected by (a sort of) authorship's right.

[8] Interestingly, in some Latin languages such as Italian, (and Spanish and French) the term 'Intellectual property right' is translated literally as "protection of authorship's rights" (protezione dei diritti di autore), with no reference to 'property'. I am sure that colleagues in Law and Ethics would be able to step in this debate much more elegantly than I can ever do here.

## Redistributing value within the “Data as Intellectual Property Right” framework

To the extent that individual data is used (collected, aggregated, and analysed) by the firm to increase its intangible assets, it should be treated as use of an intellectual asset and remunerated through the payment of a licence fee. This would change the nature of the contract between individual and company: rather than being paid a wage within a labour contract, the individual would be paid for the use concession of a licensable asset, her authorship’s right. It would also align better with the nature of a balanced exchange: intellectual property is owned by the worker or consumer and used by the firm, who pays a licence fee to do so.

By considering it as a licensable asset, data generators could choose to be paid a license use fee when data analytics are used for private purposes and feed into profits (e.g. marketing analytics). Alternatively, they can choose to openly share private information in case personal data feeds into public knowledge (e.g. health)

There are some advantages of an approach based on “Data as Intellectual Property Right” and a remuneration of data as a licensable asset over the others. For example, it could (i) Reduce the infrastructural burden of administering a digital tax or changing digital ownership; (ii) Ensure dismissed workers do not lose their rights on data ownership once they are out of the labour contract; (iii) Reduce the likelihood that certain workers miss being paid a wage against the use of their data; (iv) Ensure that firms keep paying an IPR to consumers who have completed / exhausted their consumption transactions, but who have provided data that continues to contribute to the intangible assets of the firm; (v) This way we do not necessarily tax innovative firms, but redistribute profits directly.

This proposal would require first and foremost a change in mindset about the nature and the value of ‘big data’ in a context in which – as it has been put forward – data is the ‘new oil’[9]. Indeed, data are not a source of rent (as oil is) for those that stock it, but a source of profit.

[9] See The world's most valuable resource is no longer oil, but data. The Economist, May 2017. See also If data are the 'new oil', how can their value be shared fairly? Policy@Sussex, 11 April 2019

Data ownership is arguably the most diffused source of value and should first be tracked and then recognised at time of provision.

This is clearly only a first step. The technicalities and their design should be the object of specific further analysis. For instance, implementation would require a definition of the duration and the dimension of the licence fee, and the tax regime that should be applicable (or not) to it. One could think of a tax-free licence to be perceived by data subjects, which would align to the redistribution rationale put forward here. However, this could entail a perverse incentive to generate data to maximise financial compensation, similarly to what observed for the "Data as Labour" framework.

Underpinning all of the frameworks above is an assumption of **traceability** of the data generated by individuals. Redistribution of data value would need to start as soon as the individual data are collected in some aggregate form for further use. Traceability is therefore the single, most crucial implementation challenge that affects all three frameworks. Arguably, the GDPR and related fundamental data protection regulations are a milestone to build upon to implement traceability**. Any record of individual consent to use of their data as a result of GDPR compliance is potentially a traceability starting point that allows implementing all the forms of data value redistribution illustrated above**. This can become a General Data Tracking Regulation that would allow corporate actor to be accountable for any profitable use of individual data.

## 4. Challenges: The notions of data property, ownership and right

The founding rationale of the GDPR is that personal data protection is a fundamental right. This gives the EU a pioneering advantage in the field of data regulation, in comparison with the absence of any similar regulation in the US, Japan or China. It should be borne in mind that the GDPR also regulates and encourage the "free movement" of personal data within the EU.

For the purpose of this note, that is mainly finding a feasible ground to redistribute the value of data, the rationale of the GDPR might pose a number of challenges.

If personal data protection is a fundamental right, ensured to each and every citizen (regardless of their 'worker' or 'consumer' status), would the consideration of an individual's data as their intellectual property undermine its nature as a fundamental right? In other words, would the monetisation of personal data – in any of the forms above, albeit beneficially - be inconsistent with the nature of fundamental right?

Are the *right to privacy protection* and the *right to be compensated as 'author'* compatible? Would the creation of a market for data be voiding or conflicting with the right to individual data protection and the regulation of free data movement?

Have research and practice tackled these issues?

There have been numerous debates and local experiments around the **democratisation of data governance,** data **sovereignty.** Institutions such as the Open Data Institute are questioning the **appropriateness of the concept of data ownership** and the natural monopoly that data owners – either private or public – represent. Further considerations on the role of **encryption** versus data sharing deserve more research efforts and we do not enter this debate here. However, with in mind the aim grounding a policy to redistribute data value, we develop a few considerations around the challenges above, based on further scholarship.

It has been proposed that a Data Trust law could be implemented in addition to the GDPR (Delacroix and Lawrence, 2018). Trust law is based on the "duty of care" of the actor that (gathers), treats and uses data. Delacroix and Lawrence (2018) argue for the implementation of a variety of Data Trusts, each instantiating a particular way of balancing out risks and responsibilities. Such an ecosystem of Data Trusts would allow individuals to switch Trusts when needed or preferred. Its success would depend on the Trusts' ability to implement their functions around the rights on data portability and data erasure (e.g. "right to be forgotten", arts 20 and 17 of GDPR).

**Data Trusts could therefore circumnavigate the issues around 'data property' by focusing on 'data rights'**. It is undoubtable that different data would require different rights, from full portability, to access and erasure. A Data Trustee would need a mandate to exercise rights on behalf of data subjects (Arts 80(1) of GDPR currently envisage such mandate only in relation to arts 77-79) (Delacroix and Lawrence, 2018). By facilitating access to 'pre-authorised' aggregated data, instituting a Trust may remove key obstacles to the untapped research potential relying on 'Big Data' (see also Corrado, 2019).

However, the take-off of Data Trusts might be constrained by the same implementation challenges mentioned above in the case of MIDs within the "Data as Labour" framework. First, individuals might have minimal incentives to register and/or having the necessary competences or willingness to identify the ones they wish to register with, in a context of a multitude of Trusts. Also, shared provenance is more common than is assumed, and could prove problematic to the implementation of a Data Trust (e.g. genetic data, ambient surveillance, social media feeds), as the authors also mention (Delacroix and Lawrence, 2018). More in general, a full, scaled up adoption of Data Trust is difficult to predict, particularly in a context of trust competition as the one advocated by the Delacroix and Lawrence (2018).

## 5. Considerations for policy and further research

The rationale behind the GDPR might not necessarily be incompatible with the frameworks above. Within the "Data as Capital" framework (as in many constitutions), we could consider the freedom to start an economic activity as a fundamental right. Similarly, within the "Data as Labour" framework, the right to a decent job is, in many cases, protected as a fundamental right. Within the "Data as Intellectual Property Right" framework, both private property and intellectual property rights are contractual rather than fundamental rights. Even so, data use by non-owners might legitimately require their owner to be compensated, and the non-owner to abide to specific duties. For instance, in the case of (data) trust law, the **trustee** having the assets (data) holds a right (over data use) which is subject to a duty owed to the specific **beneficiary** (the individual providing their data, or data subject) or to be used exclusively for **charitable purposes (for instance, public benefit).**

The challenges considered above, albeit crucial to steer the debate on data governance, seem to overlook the fact that the notion of 'data property' is already a reality, and has been for a long time.

The proposal of treating data as an intellectual property outlined above needs, for the moment, to build upon the existing data governance structure and its unintended consequences in terms of value redistribution, regardless of whether the concept of data ownership is appropriate or not. To this purpose, we need to frame the debate on the basis of a series of common definitions, and a common understanding of what data value is and new method to track it. Also, the appropriate framework(s) should not only assess whether we need a process of redistribution, but also identify who is in the relevant position to implement it.

The IPR proposal not only does not necessarily risk voiding the rationale behind GDPR, privacy and data protectionism, but could usefully build on it. For instance, the GDPR and other EU regulations can play a fundamental role in implementing a fairer redistribution of data value, as they provide an **unprecedented awareness to individuals to consent, deny** or **limit** the use of data and **choose the recipient.** Also**,** any record of individual consent to the use of personal data as a result of GDPR compliance is a **potential traceability starting point.** Tracing means providing a channel to **identify actors and stages of data treatment** and measure the value of data analytics.

The pleas for data trust are valuable, and it is worth continuing the debate. However, when the issue of scalability occurs, we could think of a **European Data Trust,** which is not mutually exclusive with local, small scale experiments of data commons, it could actually be a governance benchmark.

More research for action is needed to explore all these issues in further depth, particularly those that underpin reflections on data value redistribution, and I currently propose to prioritise examining the following:

1. Is the notion of Individual Data Right helpful at all? Does the GDPR protect the Data Analytics segment of the Data (Global) Value Chain, rather than the “raw” material only (the individual data) and if not, should it? From a data value chain perspective, this means protecting data intermediates alongside raw data. This relies on the assumption, mentioned above, of an effective infrastructure for data traceability to be in place. We would need to frame a new Governance of the

Commons rather than the individual and the GDPR could become a global standard. Here is a matter of **scale of the data treatment** rather than the protection of the individual data as a fundamental right.

2. Where should the enforcement of data protection start? Where is the power in the current structure of data governance? Again, the crucial issue might be to regulate along the whole data value chain, from individual protection, through the (private) data collection infrastructure, therefore enforcing transparency and accountability of actors that own the infrastructure. **This represents a necessary (but not sufficient) condition to implement any of frameworks put forward in this note.**

3. To what extent it is feasible to use tax gains to create a European data public Trust, a public mediator of individual data, assuming this is the right institutional level to protect intermediate data markets? To what extent the current examples of local implementation of Data Trust or commons **are useful to test the feasibility of a European level Trust**? What are the bottlenecks, the financial and use constraints to scale up local experiences?

## References

Bénabou, Roland & Jean Tirole, 2006. Incentives and Prosocial Behaviour. *The American Economic Review* 96(5), pp. 1652-78.

Erik Brynjolfsson & Daniel Rock & Chad Syverson, 2018. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, NBER Chapters, in: The Economics of Artificial Intelligence: An Agenda, pages 23-57 National Bureau of Economic Research, Inc.

Buchanan, James M., 1965. An Economic Theory of Clubs, in *Economica*, New Series, Vol. 32, No. 125, pp. 1-14.

Corrado, Carol & Charles Hulten & Daniel Sichel, 2009. Intangible Capital And U.S. Economic Growth, *Review of Income and Wealth, International Association for Research in Income and Wealth*, vol. 55(3), pages 661-685, September.

Corrado, Carol, 2019. *Data as an Asset: Expanding the Intangible Framework*. Presented at the EMAEE 2019 Conference on the Economics, Governance and Management of AI, Robots and Digital Transformation – Held at SPRU, University of Sussex, 3-5 June 2019. http://www.sussex.ac.uk/spru/newsandevents/events/emaee2019.

Delacroix, Sylvie and Lawrence, Neil, 2018. Disturbing the 'One Size Fits All' Approach to Data Governance: Bottom-Up Data Trusts" (October 12, 2018). Available at SSRN: https://ssrn.com/abstract=3265315 or http://dx.doi.org/10.2139/ssrn.3265315

Foray, Dominique, 2004. *The Economics of Knowledge*. MIT Press.

Ibarra, Imanol Arrieta & Goff, Leonard & Hernández, Diego Jiménez & Lanier, Jaron & Weyl, Glen, E., 2018. Should We Treat Data as Labor? Moving Beyond "Free", *American Economic Association Papers & Proceedings*, 2018, Vol. 1, No. 1, https://ssrn.com/abstract=3093683.

Jones, Charles I. & Tonetti, Christopher. 2019. Non-Rivalry and the Economics of Data, https://www.nber.org/papers/w26260.

Morone, Jennifer, 2019. *Data as Labour for Data Dignity*, Presented at the EMAEE 2019 Conference on the Economics, Governance and Management of AI, Robots and Digital Transformation – Held at SPRU, University of Sussex, 3-5 June 2019. http://www.sussex.ac.uk/spru/newsandevents/events/emaee2019.

OECD, Oslo Manual 2018 – Guidelines for collecting, reporting and using data on innovation, 4th edition, October 2018, http://www.oecd.org/science/oslo-manual-2018-9789264304604-en.htm

Posner, Eric & Weyl, Glen E., 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.

Savona, Maria, 2019. If Data are the 'New Oil', how can their value be shared fairly? Policy@Sussex, 11 April 2019.

Soete, Luc & Kamp, K, 1997. Taxing consumption in the electronic age, *Intermedia* 25(4), August 1997, pp 19-22.

# Recent papers in the SPRU Working Paper Series:

**September**

2019-20. Teaming up with Large R&D Investors: Good or Bad for Knowledge Production and Diffusion? *Sara Amoroso and Simone Vannuccini*

2019-19. Experimental Innovation Policy. *Albert Bravo-Biosca*

**August**

2019-18. Relating Financial Systems to Sustainability Transitions: Challenges, Demands and Dimensions. *Chantal P. Naidoo*

2019-17. Innovation and Self-Employment. *Tommaso Ciarli, Mattia Di Ubaldo and Maria Savona*

2019-16. Integration in Global Value Chains and Employment in Europe. *Filippo Bontadini, Rinaldo Evangelista, Valentina Meliciani and Maria Savona*

2019-15. Revisiting the Natural Resource 'Curse' in the Context of Trade in Value Added: Enclave or High-development Backward Linkages? *Filippo Bontadini and Maria Savona*

**July**

2019-14. The Impact of Increasing Returns on Knowledge and Big Data: From Adam Smith and Allyn Young to the Age of Machine Learning and Digital Platforms. *Yao-Su Hu*

2019-13. Eco-Innovation and Firm Growth in the Circular Economy: Evidence from European SMEs. *Pelin Demirel and Gamze Ozturk Danisman*

**Suggested citation:**
Maria Savona (2019). The Value of Data: Towards a Framework to Redistribute It. SPRU Working Paper Series (SWPS), 2019-21: 1-22. ISSN 2057-6668. Available at: www.sussex.ac.uk/spru/swps2019-21

Science Policy Research Unit
University of Sussex, Falmer
Brighton BN1 9SL
United Kingdom

SPRU website: **www.sussex.ac.uk/business-school/spru**
SWPS website: **www.sussex.ac.uk/business-school/spru/research/swps**
Twitter: **@spru**

UNIVERSITY OF SUSSEX | BUSINESS SCHOOL