

Policy Brief

POLICY@SUSSEX | MARCH 2020

Governance of Data Value



EXECUTIVE SUMMARY

Interest in personal data has been growing unprecedentedly. Issues of privacy and power are at the forefront of policy debates. Yet, these concerns seem to overlook the issues of **concentration of equity value** (stemming from data value, henceforth used interchangeably) that underpins the current structure of big tech business models.

Economists have failed to predict the **massive concentration of data value** in the hands of large platforms and underestimated the complexity of the **political economy** of data value concentration.

A systematisation of recent research leads Professor Savona to propose a **novel data rights approach**, that redistributes data value to achieve economic justice whilst not undermining the range of ethical, legal and governance challenges that this poses. By granting **authorship rights** to data generators and enforcing large platforms – as large publishers – to remunerate them, should they choose to be remunerated, it would be possible to start tackling data value redistribution and increase individual agency over personal data.

The approach requires designing a **novel institutional architecture for data value governance**, that creates synergies between decentralised and centralised governance models and maximises the public use and value of data.

Key findings

- Personal data is closest to a club good. Data is an intangible yet durable asset as it does not become obsolete. Its value is in its scale.
- The business models of big tech rely on a complex (and opaque) integration of layers.
- Different governance models depend on what rights data require and for which purposes.
- Governance of data value includes enforcing data taxation systems and recognizing and protecting authorship's right to data generators
- The approach requires designing a **novel institutional architecture for data value governance**, that creates synergies between decentralized and centralized systems and maximizes the public use and value of data.

MARIA SAVONA

Professor of Innovation and Evolutionary Economics

Email: M.Savona@sussex.ac.uk

Twitter: @maria_savona



US

UNIVERSITY
OF SUSSEX

The state of the art in governing data

There is a growing awareness of the threat of violations of privacy and power abuse, sparked by practices of electoral behaviour manipulation unveiled in the Cambridge Analytica scandal (Privacy International, 2019, among other analyses). Yet, market and power structures, dominated by the few well-known big techs, have gone beyond the forward-looking views laid down in the popular book *Surveillance Capitalism* by Shoshana Zuboff (2019), and in contributions by other prior insiders, such as Roger McNamee, as narrated by Barth (2019). These works compellingly reveal how citizens are poorly aware of how fine-grained the massive amount of data they generate is, and how data is gathered, stocked, treated and analysed, predominantly by big tech. Whilst these quasi-monopolies own the *digital infrastructure* to do so, they do not own the individual data that provide the raw material for data analytics.

The European Commission (EC) has been at the forefront of global action to promote convergence of the governance of data (privacy) within the EU, including, but not limited to, the well-known EU GDPR (General Data Protection Regulation) (EU, 2016), enforced in May 2018, and more recently, the EC White Paper On Artificial Intelligence – A European approach to excellence and trust (2020) and the EC “A European Strategy for Data” (February 2020). Some awareness has increased at least among (EU) citizens, who can now decide to actively provide, deny or withdraw consent to the use of their data by a specific actor, or exercise their right to be forgotten, while companies find that complying with the GDPR is, at best, a useless nuisance.

Databases have been accounted for as companies’ ‘innovation investment’ since the beginning of the 1990s. The first edition of the Oslo Manual, the OECD international guideline to collecting and using data on innovation in firms, dates back to 1992 and originally included the collection of databases on employee best practices as innovation expenditures (see also the Oslo Manual’s latest edition: OECD, 2018). More recently, data is measured as firms’ ‘intangible asset’ (Corrado et al., 2009 among the pioneers).

What has changed over the last decade? *The scale of data generation* has reached a dimension that its management and control might have already gone well beyond the capacity of the very tech giants we are all feeding. Concerns around data governance, data rights, data privacy, and a sense of imminent impingement of our democracies, have been raised, though they *might be too little and too late* to lead to effective and timely action. One of the reasons why it is so, is that they seem to overlook the issue of concentration of equity value that underpins the current structure of big tech power. Here I argue that economists have failed twice and propose a potential way out.

A political economy of data: the basics

The first failure of economists has been the underestimation of the complexity of the political economy of data, despite some attempts (Posner and Weyl, 2018; Bennett Institute, 2020).

Arguably, the political economy of data builds – among other aspects - upon the economic nature of data in terms of *rivalry* and *excludability*.

A *private good* is excludable (an individual can be denied access to it) and rivalrous (access and consumption by an individual make it unusable for another individual). Property rights are based on the notion of excludability and rivalry and are usually contractual rights recognised by the state.

A *public good* is not excludable nor rivalrous. The creation of public goods is deliberately pursued in the collective interest. An example of a public good is *public knowledge*, although it relies on the accumulation of *information*, which is usually a *private good*. Granting Intellectual Property Rights (IPR) aims at protecting and remunerating *intellectual creation* to preserve individual incentives to contribute to a public interest, rather than the exclusive fruition of a private good. Intellectual property rights are still contractual rights, though they are enforced to preserve the creation of public knowledge.

A *common good* is rivalrous but not excludable (typically an extractive resource, whose use is limited). A club good is excludable but not rivalrous, so that an individual can be denied/deny access to and use of it, though once this access is allowed, the good is non-rivalrous. It does not become used-up and its use can be enjoyed by multiple individuals, as it is *reproducible*.

The *conundrum* around personal data is that it is closest to a club good. Data is (potentially) highly excludable but non-rivalrous, as its reproducibility and fruition are virtually infinite, at a zero-marginal cost. Data is an intangible yet durable asset as it does not become obsolete. Its value is in its scale. Yet, unlike traditional mass-production goods, data is far from being standardised and homogeneous, rather, personal data is a unique combination of nested individual characteristics.

It is not straightforward to establish what kind of rights data requires and for which purposes. For instance, data protection laws, as enacted by the EC, assume privacy to be a fundamental right, embedded in the European Convention on Human Rights. The GDPR seems to regulate the rights to excludability (privacy) comparatively more than the management of rivalry (e.g. use by third parties, right to be forgotten). However, identifying the rights that regulate the distribution of data value (private or public) and taxation is not straightforward, and to my knowledge there has not yet been attempts to design (let alone enforce) such rights.

GOVERNANCE MODELS FOR REDISTRIBUTING DATA VALUE AND IMPLEMENTATION CHALLENGES

Unpacking data value chain

That data is part of the intangible capital of firms is not at all new and it has been largely unquestioned, until now. Data has long been considered as firms' 'knowledge-based capital' or 'intangible assets', and the measurement and economic impact of intangibles have been the object of an established branch of literature (Corrado et al., 2009 among the pioneers). However, economists have overlooked the evolving nature of the data value chains, missed the opportunity to identify, let alone to quantify, both its final product and the sources of its (re)production.

The business models of big tech rely on a complex (and opaque) integration of layers. This includes data gathering, accumulation and in-house treatment; third parties, intermediate users and providers of data analytics; as well as interfaces that offer 'free online services' to individuals. We do not know whether the (equity) *value* of platforms is truly aligned with the scale of data *accumulation*, to stick with the capital metaphor, or to something else. Intangible assets include, besides data, investments in R&D, patents and licenses, trademarks, organisational capital, training, engineering, design and so on (see Oslo Manual's latest edition, OECD, 2018). However, as most of the equity value from data analytics is in advertising, it is difficult to argue that all intangibles are *knowledge-based capital*.

Economists have *de facto* legitimated the notion of data as intangible capital, though not fully kept up with the understanding of how the data value chain was evolving until it was (too) late (for instance Brynjolfsson et al., 2018; Brynjolfsson et al., 2019) to prevent the current quasi-monopolistic market structure of large platforms.

Taxing intangible capital (more)?

Data has long been considered as an intangible asset of companies and is included in firm balance sheets. One could therefore argue that it would be straightforward to design and implement an effective tax system on intangible capital by a (supranational?) fiscal authority.

A seminal proposal of a taxation system in the form of a 'bit tax' was put forward some twenty years ago by Soete and Kamp (1997), who first advocated the taxation of the number of 'bits' rather than of the value added of intangibles, which was then at an embryonic stage.

An adequate system of taxation of current large platforms could be designed in a similar way, by taxing large platforms at the start for each individual's data collected. More recently, in the UK there is a debated proposal on introducing a new Digital Service Tax (HRMC, 2019).

The implementation of a 'data tax' is, however, based on the heroic assumption that it is straightforward to track data along its value chain. This would entail an appropriate price system to quantify costs of storage, aggregation and treatment. Currently, there is no price system to estimate the value data. Balance sheets at best measure data analytics similarly to how software investments are estimated, that is in terms of work compensation of data scientists and engineers employed in data analytics tasks, or indeed as 'purchased software' (Corrado, 2019). In addition, the very nature of data would make the role of supranational fiscal institutions more appropriate, yet, in the context of increasingly undermined traditional national tax bases, this opens up a Pandora's box of implementation challenges, all the more so outside the EU.

Creating data labour markets or radicalising capitalism?

It has been argued (Posner and Weyl, 2018) that as the "powerhouse of the digital economy", the few big tech companies exploit the lack of public understanding of how individual data is collected and treated, there is a missing labour market for data generators. The notion of labour dignity becomes 'data dignity'. For instance, Machine Learning (ML) trainers could be remunerated for generating the high-quality data that helps feed into AI. There will be a market for ML Collective Intelligence that remunerates the sources of its (re)production.

This view offers the opportunity to revisit the traditional forms of collective bargaining and collective representations and adapt them to the digital ecosystem. The Mediators of Individual Data proposed by Posner and Weyl could allow for (i) collective bargaining with big tech; (ii) quality certification of data.

Creating such a credible institutional actor that can represent and collectively bargain on behalf of data labourers is a necessary, though not sufficient, condition to make this governance model work. People might have intrinsic and extrinsic incentives to generate data as a job (Bénabou and Tirole, 2006). Altruistic incentives would increase the likelihood of generating high-quality data and a sense of belonging to a community. However, perverse incentives might lead people to generate a mass of low-quality data to maximise financial remuneration. As compellingly argued by Pavel (2019), this might well be the case, when more vulnerable, less skilled workers have perverse incentives to generate data as a result of income constraints. However, if they are less educated and skilled, their low-quality data might be remunerated less, creating a vicious circle. Current labour markets issues, such as technological unemployment, skill-biased technical change, and other forms of inequality would just be reproduced in a data labour market.

Advocating for data labour markets to address data value redistribution is an endeavour whose success is linked to an adequate system of collective representation and bargaining. By its very nature, it risks reproducing – and possibly exacerbating – the labour markets' inequality which, driven by technical change, leaves behind unskilled and precarious workers.

LARGE PLATFORMS AS LARGE PUBLISHERS: RECOGNISING AUTHORSHIP RIGHTS TO DATA GENERATORS

This section explores the basics of a rationale to consider personal data as an intellectual creation, and recognise *authorship rights* to the individual who has generated it.

First, personal data makes the (digital) identity of an individual. Hence, the concepts of *data ownership* and *property* (i.e. an individual owning her data) can be argued to be fairly meaningless, as the individual is at once the original intellectual creation, embedded in their own data. Personal data results from the complex set of individual histories, knowledge, preferences and value systems. Even allowing for identity to be a social intersection (Immorlica et al., 2019) that is, based on personal data from and shared with others, it is down to the individual to consent to and allow the use of their identity. This is a unique creation, worthy of protection, recognition and remuneration, in case of reproduction, aggregation and treatment processes by any third party. In summary, because of its nature as a club good and the uniqueness of each individual data, I argue that an individual might claim authorship rights over their data.

Second, within this proposed governance model, big tech companies and linked third parties that collect, reproduce, analyse, extract value from individual data and increase their intangible assets, are to be considered *publishers*, rather than *platforms*. They should be required and

ideally be enforced to recognise, protect and remunerate individual authorship rights, for life, and regardless data generators' job status. Big tech are extracting value from individual intellectual creations similarly to how they use other companies' IPR against payment of a licence fee.

Third, being entitled to authorship rights would potentially increase individual agency over own data and give content to the notion of data dignity. Data subjects could choose to be paid a use license fee when data analytics are used for private purposes and feed big techs' profits (e.g. marketing analytics). Alternatively, they can *choose* to openly share private information in the case where personal data feeds into public knowledge (e.g. research).

There are some advantages of an approach based on recognising, protecting and remunerating authorship's rights over the other governance models. 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 wages once they are out of the labour contract; (iii) Ensure that large platforms keep paying authorship rights to consumers of digital services, who have completed/exhausted use of online services, but who have provided data that continues to contribute to the intangible assets of the firm; (iv) innovative firms are not necessarily taxed, but profits are redistributed directly; (v) Authorship rights can also be collectively licensed, for limited purposes.

A bold vision to design a governance model based on data rights that pursues economic fairness alongside social justice requires rethinking the relationship between centralised and decentralised data governance, as hinted at in Pavel (2019). A seemingly successful example of an optimal synergy between centralised and de-centralised data governance, when individuals choose free data sharing when it serves a public purpose, is Taiwan. At exceptional times of public health emergency such as the present Covid-19 pandemic, Taiwan seems to have contained the spread of contagion better than other countries. The country has managed to digitally track individuals with symptoms and their contacts to contain the spread, with a mix of community based app, releasing data on symptoms and positions, and a fast reaction by the digital minister who has coordinated the government's response (Lanier and Weyl, 2020). This synergy is most likely facilitated by an exceptionally high digital literacy of Taiwanese population (e.g. of elderly people) and a shared sense of public purpose. A crucial question is however the "reversibility" of these tracking apps when the situation gets back to normal and people would have shared a high amount of sensible personal data.

POLICY IMPLICATIONS

- This note is a very first step to ground a multidisciplinary and wide-ranging reflection on the complex issue of data governance, that addresses economic justice, individual contributions to public value, alongside privacy issues. The challenge of redistributing data value is often not tackled by those mainly interested in data rights. Similarly, those (few, so far) interested in tackling value redistribution are often oblivious to the complexity of the ethical and legal boundaries of data rights.
- The governance models revisited and proposed here can be summarised in the table below, with main pros and cons in terms of implementation.
- The data right approach proposed here would not be incompatible with the idea of data dignity. This should be based on deliberate actions to increase individual awareness and agency over data. Individuals should be empowered and given the choice of belonging to a community and contributing to social good, or having their rights remunerated.
- A bold vision to design a governance model based on data rights that pursues economic fairness alongside social justice requires rethinking the relationship between centralised and decentralised data governance, as hinted at in Pavel (2019). A competitive system of several bottom-up and purpose-specific data trusts as recently advocated (Delacroix and Lawrence, 2018) would need to be complemented by centralised public institutions that regulate them, ensure scalability and enforce compliance by big tech. The EC has the comparative advantage of a first mover as a regulator of the data system. We should start from there.

Categorisation	Data as capital	Data as labour	Data as authorship right
Governance	Taxation system	Wage	Licence fee
Pros	<ul style="list-style-type: none"> • Data taxed at outset • Straightforward to implement through fiscal authority 	<ul style="list-style-type: none"> • Balanced distribution of data value including collective bargain with big tech and quality certification • Altruistic incentives may increase quality of data generation 	<ul style="list-style-type: none"> • Burden of administering a digital tax or changing digital ownership reduced; • Dismissed workers do not lose their rights once they are out of the labour contract; • Firms pay authorship rights to consumers who have completed use of online services, but whose data continues to contribute to the intangible assets of the firm; • Authorship rights can be collectively licensed
Implementation challenges	<ul style="list-style-type: none"> • Data value chain as a black box • No price system to value data • Governments may be reluctant to impose digital taxes on 'innovative' firms • National fiscal authorities should coordinate 	<ul style="list-style-type: none"> • Perverse incentives may lead to mass generation of low-quality data • Risk reproducing and exacerbating labour market inequalities • Credible institutional actor required 	<ul style="list-style-type: none"> • Need novel institutional architecture that coordinates centralised and decentralised governance systems • Credible institutional actor required

Table: Data Governance Models

REFERENCES

- Barth, B., 2019. [Big Tech's Big Defector](#). The New Yorker, 25 November 2019
- Bénabou, Roland & Jean Tirole, 2006. Incentives and Prosocial Behaviour. *The American Economic Review* 96(5), pp. 1652-78
- Bennett Institute for Public Policy, 2020. The Value of Data. Policy Implications. February 2020. <https://www.bennettinstitute.cam.ac.uk/publications/value-data-policy-implications/>
- Brynjolfsson, Erik & 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.
- Brynjolfsson, Erik & Avinash Collis & W. Erwin Diewert & Felix Eggers & Kevin J. Fox. 2019. GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy. National Bureau of Economic Research Working Paper 25695, <http://www.nber.org/papers/w25695>
- 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/newsanevents/events/emaee2019>.
- Delacroix, Sylvie and Lawrence, Neil, 2018. Disturbing the 'One Size Fits All' Approach to Data Governance: Bottom-Up Data Trusts. <https://ssrn.com/abstract=3265315> or <http://dx.doi.org/10.2139/ssrn.3265315>
- European Union, 2016. Regulation 2016//679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR)
- HRMC, 2019. Introduction of the new Digital Services Tax. July 2019, <https://www.gov.uk/government/publications/introduction-of-the-new-digital-services-tax/introduction-of-the-new-digital-services-tax>.
- Immorlica, Nicole & Jackson, Matthew O. & Weyl, Eric Glen, 2019. Verifying Identity as a Social Intersection. Available at SSRN: <https://ssrn.com/abstract=3375436> or <http://dx.doi.org/10.2139/ssrn.3375436>
- Lanier, Jaron & Weyl, Glen E., 2020. How Civic Technology Can Help Stop a Pandemic. *Foreign Affairs*, March 20, 2020. <https://www.foreignaffairs.com/articles/asia/2020-03-20/how-civic-technology-can-help-stop-pandemic>
- 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>
- Pavel, Valentina, 2019. Our Data Future. *Privacy International Long Read*, July 2019. <https://privacyinternational.org/long-read/3088/our-data-future>
- Posner, Eric & Weyl, Glen E., 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press
- Privacy International (2019). Cambridge Analytica, GDPR - 1 year on - a lot of words and some action. <https://privacyinternational.org/news-analysis/2857/cambridge-analytica-gdpr-1-year-lot-words-and-some-action>
- Savona, Maria 2019a. 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
- Savona, Maria, 2019b. 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.
- Zuboff, Shoshana, 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (1st ed.). Public Affairs. ISBN 1610395697

This briefing is supported by the Policy@Sussex initiative which connects social science research to a wide range of stakeholders.

This policy brief extends a recent [VOX EU Column on Governance models for redistribution of data value \(17 January 2020\)](#) and is based on a prior version of a research note on [The Value of Data. Towards a Framework to Redistribute It](#)