

Valuation amidst anonymisation: case studies from the individual to the commons and back again using Viljoen's axes of horizontal data relations

Georgia Meyer (London School of Economics and Political Science (LSE))

Introduction / abstract

This paper will present an empirical study of anonymised personal data across a narrow but illustrative range of cases of public and private data collection and sharing contexts. In exploring the way that conversations about anonymisation (Stalla-Bourdillon & Knight 2017), as well as the rapidly rising interest in and sophistication of PETs (McCarthy 2020), might meet the growing body of work on the relational nature of data (Coyle 2022; Viljoen 2021), this study hopes to provide a set of associated concepts to accompany Viljoen's notion of horizontal data relations. This is with a view to further establishing the theoretical dexterity required to assess the multi-dimensional nature of data's (social and economic) value*.

Literature overview

Some of the most ground-breaking work in understanding the nature and value of data is tackling the ontological assumptions about data as an object (that can be quantified, measured and owned) versus data as a temporally emergent phenomena that is inherently relational and that has considerable, albeit hard to measure, societal benefits (Ramesh & Kapoor 2020). These worldviews about the nature of data are both very much alive and exist side by side. Yet current technical, legal and economic lexicons that seek to identify (and create) 'its' value and protect individual rights to privacy are conceived entirely within the former conception of data.

This picture is further complicated by two further considerations: (i) current dominant paradigms for data protection law assert the anonymised personal data is not subject to the provisions of data protection legislation (DPDI 2022); (ii) the rise in sophistication and deployment of PETs suggesting the possibility for a more widespread, varied and intricate, use of anonymised personal data. In the case of the former questions arise about the absence of consent mechanisms at the level of the individual, in the latter case questions arise about the nature of the valuation processes applied to anonymised personal data.

This study will take as its starting point that anonymised personal data is outside the remit of data protection legislation and apply Viljoen's (2021) concept of horizontal data relations to explore the social dimensions of data valuation across a range of contexts. It will begin with a literature review in three parts: (i) studies that measure / discuss the value of personal data in terms of an asset that is bounded and can be delineated to single

identifiable individuals; (ii) studies that discuss the nature / value of personal data in terms of its relational attributes; (iii) literature that discusses the saliency of concepts like anonymisation and distinctions between personal and non-personal data. Oftentimes literature on social benefits (or costs) to data are discussed using economist's notions of externalities. But as Viljoen points out, this casts data's multidimensional value generating nature as an incidental rather than the instrumental purpose for its collection, processing and application. This study seeks to further embed this observation by foregrounding the notion anonymised personal data which shifts the conversation away from the debate about individual informational privacy and explicitly onto data relations in aggregate anonymised form.

This three-part literature review will be used to develop a set of concepts that can accompany Viljoen's (2021) notion of horizontal data relations and be applied to data valuation chains that will form the main body of the study. These concepts will be developed to: (i) unpack the types of categorisation processes that are applied to anonymised personal data in order to establish population-level insights thereby enacting horizontal data relations, and, (ii) provide some conceptual scaffolding around how social value from data is conceived and measured.

Proposed Methodology

The empirical work of this study will involve applying the set of accompanying concepts to Viljoen's (2021) notion of horizontal data relations to a range of anonymised personal data value chains. The methodology will comprise a review of cases of generating value from anonymised personal data that sit across a variety of contexts, specifically: healthcare patient data, search history data, and IoT data.

Potential contribution

Provide some routes to expanding the notion of horizontal data relations with a view to both standardising its application and increasing its contextual dexterity.

References

Coyle, D. (2022). Socializing Data. *Daedalus*, 151(2), 348–359.

https://doi.org/10.1162/daed_a_01921

Data Protection and Digital Information Bill, (2022) (testimony of DCMS)

<https://publications.parliament.uk/pa/bills/cbill/58-03/0143/220143.pdf>

McCarthy, N., & Fourniol, F. (2020). The role of technology in governance: The example of Privacy Enhancing Technologies. *Data & Policy*, 2, e8. <https://doi.org/10.1017/dap.2020.8>

Ramesh, A., & Kapoor, A. (2020). Principles for Revenue Models of Data Stewardship. Aapti Institute. <https://thedataeconomylab.com/2020/07/31/principles-for-revenue-models-of-data-stewardship/>

Shkabatur, Jennifer, The Global Commons of Data (October 9, 2018). Stanford Technology Law Review, Vol. 22, 2019, GigaNet: Global Internet Governance Academic Network, Annual Symposium 2018, Available at SSRN: <https://ssrn.com/abstract=3263466>

Spiekermann, K., Slavny, A., Axelsen, D. V., & Lawford-Smith, H. (2021). Big Data Justice: A Case for Regulating the Global Information Commons. *The Journal of Politics*, 83(2), 577–588. <https://doi.org/10.1086/709862>

Stalla-Bourdillon, S., & Knight, A. (n.d.). Anonymous data v personal data - a false debate: An EU perspective on anonymization, pseudonymization and personal data

Viljoen, S. (2021). A Relational Theory of Data Governance. *The Yale Law Journal*, 82. *and risks, though societal risks are not an explicit part of the study