GLOSSARY ENTRY

OPEN ACCESS

PEER REVIEWED

Data intermediary

Heleen Janssen *University of Amsterdam*
Jatinder Singh *University of Cambridge*

Abstract: Data intermediaries serve as a mediator between those who wish to make their data available, and those who seek to leverage that data. The intermediary works to govern the data in specific ways, and provides some degree of confidence regarding how the data will be used.

> This article belongs to the **Glossary of decentralised technosocial systems**, a special section of *Internet Policy Review*.

Definition

A data intermediary serves as a mediator between those who wish to make their data available, and those who seek to leverage that data. The intermediary works to govern the data in specific ways, and provides some degree of confidence regarding how the data will be used.

Data intermediaries form part of a data processing ecosystem. This includes the *intermediary*, often an organisation (of some form), as well as two other key categories of stakeholder:[1] *data suppliers* who are those individuals, communities, or enterprises that make their data available, and *third parties* referring to those interested in using (processing) supplier data.

Context and description

The concept has emerged in the context of 'big data', and the increasing interest in data analytics and machine learning (Hardjono & Pentland, 2019; Stalla-Bourdillon et al., 2020; Micheli et al., 2021). Deep concerns however exist regarding opaque data practices, surveillance practices, and the systemic power and information asymmetries inherent to the current data processing ecosystems (Edelman, 2018), where organisations reap the value and benefit of data and its processing, rather than the people to whom the data pertains (Zuboff, 2015; Beer, 2017; Kitchin, 2017). Data intermediaries respond by attempting to help rebalance the relationships between those producing or with rights over data, and those seeking to use that data by offering an alternative approach to the data processing.

The data intermediary is a nascent, yet emerging concept, with the terminology still *in flux*. An intermediary's role, operation and the actions it will undertake, as well as its governance and incentive structures are very context sensitive. That is, how data intermediaries form and operate, largely depends on their purposes, the nature of suppliers and third parties they engage with, the intermediary's relationships with the suppliers and third parties involved, the data used, the means used to operate the intermediary (and whether these require a technical expertise), and so forth (see Terminologies below).

---

1. Note this is the terminology that we use; in this space, the terminology tends to vary.

Intermediaries can be proposed for a range of purposes and relationships, including by non-profits (for instance a data trust), private organisations (for instance data marketplaces), or public institutions (for instance in contexts where the public sector seeks to share data with businesses). Their business model, incentive structures, interests and governance concerns depend on the type of organisation, the purposes it pursues and the sector where it operates. A charity data intermediary might receive subsidies for enabling the sharing of health data between the public and researchers for public health purposes, whereas a commercial data intermediary might perhaps ask a third-party entrance fee for engaging with the intermediary's ecosystem. Some communities may wish to pool their data to advance particular interests for that group, or for a broader common good (Hartman et al., 2021), as might for instance happen in a research knowledge commons (Wong & Henderson, 2020).

Each data intermediary typically involves data governance measures for ensuring that data is only accessed and used as/when appropriate, giving some degree of assurance, guarantee and confidence that the rights and/or other interests of the stakeholders are properly respected and maintained – all in alignment with the intermediary's aim (see 'Governance Structures' below).

Purpose and practical usages

Intermediaries have been suggested as a way to try and tackle a range of concerns. Many proposals for data intermediaries aim in some way at countering the consolidation of power given corporate data capture and data-driven business models (Delacroix & Lawrence, 2019; Blankertz, 2020; RadicalxChange).

Often discussed are intermediaries that aim at one or more of the following:

- protecting the interests and rights of data suppliers (Reed et al, 2019; Delacroix & Lawrence, 2019; Ada Lovelace 2021; GPAI/Aapti/ODI 2021);
- rebalancing power asymmetries in data exchanges, by encouraging and empowering the data suppliers to play an active role in setting the terms of data use (GPAI/Aapti/ODI 2021);
- supporting individuals in managing their data, including helping in managing consent (Crabtree et al., 2018; Data Governance Act 2020; Ada Lovelace, 2021; Centre for Data Ethics and Innovation 2021), and in exercising their data rights (Delacroix & Lawrence, 2019; Ada Lovelace, 2021);
- enabling collective bargaining power (Hardjono & Pentland, 2019; Ruhaak, 2019; Delacroix & Lawrence, 2019);
- enabling suppliers to monetise or otherwise extract value from their data

(Ng & Haddadi, 2018; ODI, 2019; Mulgan & Straub, 2019; Benthall & Goldenfein, 2021);

- allowing the pooling of data for particular aims, e.g. for research purposes (Ausloos & Veale, 2020), investigative journalism purposes (Mahieu & Ausloos, 2020) or for the broader public interest (Scassa, 2020; see also 'data altruism' - Data Governance Act 2020; Ada Lovelace, 2021); or
- enabling the sharing of public data that is made available by governments, whereby the intermediary facilitates businesses access to that data (European Data Portal, 2019).

The above represents but a few broad categories regarding intermediary aims and example contexts in which they might be used; as the concept of the data intermediary is still developing, a variety of other purposes will likely emerge.

In terms of specific applications, data intermediaries have already been suggested and/or used in the context of the sharing of public sector data (Scassa, 2020); in the pooling of data for medical research (Centre for Data Ethics and Innovation, 2021); to enforce corporate compliance with rights, including those around employment (ACDU; WorkerinfoExchange) and data (MyDataDoneRight) or to assist in identifying discriminatory practices in credit scores (OpenSchufa).

Governance structures

An intermediary's governance mechanisms are generally proposed such that they operate in such a way that they allow for an intermediary's transparent and accountable data processing towards the other stakeholders.

Proposed data governance mechanisms include those legal, such as fiduciary duties, where intermediaries are legally obliged to act in supplier interests (Edwards, 2004; O'Hara, 2019; Delacroix & Lawrence, 2019; Ada Lovelace, 2021; GPAI/Aapti/ODI 2021), and contractual mechanisms, creating environments where data is governed under agreed terms in a controlled way (Reed et al., 2019; Micheli et al., 2020; Ada Lovelace, 2021; GPAI/Aapti/ODI 2021). Technology-backed mechanisms may also be used to allow for stakeholders to manage, monitor and control how data is accessed, used, shared, or kept in a secure manner (De Montjoye et al., 2014; Crabtree et al., 2018; Janssen et al., 2020).

These legal and technical measures can, in combination, work to provide, for example, the control and audit measures to ensure that data protection rights or trade secrets are complied with, and that data is only shared or used by third parties as appropriate. Third parties, in turn, will want assurances that the data aggre-

gate shared aligns with supplier's agreements, and the law more generally.

The power structures associated with data intermediaries can vary, for example, where the intermediary holds supplier data and performs computation over that data supplier data (i.e. taking more a 'centralised approach' to data processing), or with the suppliers holding their own data, with suppliers themselves performing computation over their data, after which the results are shared, where the intermediary works to broker and coordinate such activities (a more 'decentralised' approach to data processing).

The specifics of the governance measures employed will vary depending on the nature, aims and purpose of the intermediary, and the stakeholder rights and interests involved.

Terminologies

The term 'data intermediary', while being broad and all-encompassing, is about governance in the stakeholder interest. A range of terms have been used to describe intermediaries, which often relate to their governance structure. Common examples include:

- *data trusts,* in which the intermediary will take on responsibility to steward supplier data for agreed purposes. Data trusts may be based on fiduciary duties to act in the suppliers' interests (Edwards, 2004; Hall & Pesenti, 2017; O'Hara, 2019; Delacroix & Lawrence, 2019; GPAI/Aapti/ODI 2021), and/or be based on a contractual or statutory legal obligations (ODI, 2018; Reed et al., 2019; Ada Lovelace, 2021; GPAI/Aapti/ODI 2021);
- *data commons,* with members voluntarily 'pooling' their data for the benefit of a specific community (Wong & Henderson, 2020; Hartman et al. 2020), or for the general public interest Data Governance Act;
- *data cooperatives,* often referring to a data intermediary owned and democratically controlled by its members who delegate control over data about them (Hartman et al., 2020);
- *data collaboratives*, where participants from different sectors – including private companies, research institutions, and government agencies – can exchange data and data expertise to help solve public problems (S. Verhulst & D. Sangokoya, 2015);
- *personal information management systems* (see 'PIMS' in this glossary)*, where technology-backed systems offer data suppliers means to mediate, monitor and control how their data is accessed, used, or shared (Janssen et al., 2020);
- *data marketplaces*, *data brokers* or *trusted third parties* that work to allow the trading of data (Ng & Haddadi, 2018; Dataswift-HubofAllThings, which is

also a PIMS).

From these examples we see that data intermediaries are an emerging concept, as both the terminologies and the approaches are not only still developing, but that they may also overlap.

Debate

Ongoing discussions about data intermediaries include conversations and the development of research questions about, amongst other, how the governance structure of a data intermediary fits the purposes it pursues; whether a centralised or a decentralised approach to the data processing is appropriate for the specific intermediary's purposes, and the stakeholders involved; whether data intermediaries can, where that applies, lawfully act on behalf of the suppliers, and how such mandates relate to the supplier's rights and interests; the domains and sectors where intermediaries should be explored; the relationship between data intermediaries and personal information management systems, personal data stores and other technical infrastructures; what type of intermediary fits a certain category of suppliers (e.g. computer literate, or not), as well as questions of what robust data governance is appropriate in a specific type of data intermediary; questions of who controls and enforces the data intermediary's operations and compliance; and of who exercises oversight over the landscape with data intermediaries more broadly; and more fundamentally, whether and to what extent data intermediaries can be trusted all together.

Conclusion

Data intermediaries serve as a mediator between those who wish to make their data available, and those who seek to leverage that data. The intermediary works to govern the data in specific ways, and provide some degree of confidence regarding how the data will be used, in particular with regards to the rights and interests of those whose data is involved. Data intermediaries are a nascent, but rapidly developing concept, which lends itself for many data sharing contexts. How an intermediary operates, and the nature of its governance mechanisms, will likely depend on the specifics of the context in which it seeks to operate.

References

Ada Lovelace Institute. (2021). *Exploring legal mechanisms for data stewardship* [Report]. https://ww

w.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

Ausloos, J., & Veale, M. (2021). Researching with data rights. *Technology and Regulation*, 136-157 Pages. https://doi.org/10.26116/TECHREG.2020.010

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Benthall, S., & Goldenfein, J. (2021). Artificial intelligence and the purpose of social systems. *Proceedings of the 2021 AAAI/ACM Conference on AI Ethics and Society (AIES '21*, 1–10. https://ssrn.com/abstract=3850456

Blankertz, A. (2020). *Designing data trusts. Why we need to test consumer data trusts now* [Policy brief]. Stiftung Neue Verantwortung, Think Tank für die Gesellschaft im technologischen Wandel. https://www.stiftung-nv.de/en/publication/designing-data-trusts-why-we-need-test-consumer-data-trusts-now

Centre for Data Ethics and Innovation. (2021). *Unlocking the value of data* [Report]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004925/Data_intermediaries_-_accessible_version.pdf

Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Glover, K., Haddadi, H., Amar, Y., Mortier, R., Li, Q., Moore, J., Wang, L., Yadav, P., Zhao, J., Brown, A., Urquhart, L., & McAuley, D. (2018). Building accountability into the Internet of Things: The IoT Databox model. *Journal of Reliable Intelligent Environments*, *4*(1), 39–55. https://doi.org/10.1007/s40860-018-0054-5

de Montjoye, Y.-A., Shmueli, E., Wang, S. S., & Pentland, A. S. (2014). openPDS: Protecting the Privacy of Metadata through SafeAnswers. *PLoS ONE*, *9*(7), e98790. https://doi.org/10.1371/journal.pone.0098790

Delacroix, S., & Lawrence, N. (2019). Bottom-up data Trusts: Disturbing the "one size fits all" approach to data governance. *International Data Privacy Law*, *9*(4), 236–252. https://doi.org/10.1093/idpl/ipz014

Edelman. (2018). *Edelman trust barometer 2018—UK findings*. https://www.edelman.co.uk/research/edelman-trust-barometer-2018-uk-findings

E.D.P.B.-E.D.P.S. (2021). *Joint Opinion 3/2021 on the proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)*. https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-032021-proposal_en

Edwards, L. (2004). The problem with privacy. A modest proposal. *International Review of Law Computers & Technology*, *18*(3), 313–346. https://ssrn.com/abstract=1857536

European Commission. (2020). *Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) (COM/2020/767 final)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

European Data Portal. (2019). *Open data maturity report 2019*. European Data Portal. https://data.europa.eu/sites/default/files/open_data_maturity_report_2019.pdf

Global Partnership on Artificial Intelligence (GPAI), Aapti Institute, & Open Data Institute. (2021). *Enabling data sharing for social benefit through data trusts* [Report]. https://gpai.ai/projects/data-governance/data-trusts/enabling-data-sharing-for-social-benefit-through-data-trusts.pdf

Hall, W., & Pesenti, J. (2017). *Growing the Artificial Intelligence Industry in the UK* [Report]. UK Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy.

Hardjono, T., & Pentland, A. (2019). *Data cooperatives: Towards a foundation for decentralized personal data management*. https://doi.org/10.48550/ARXIV.1905.08819

Hartman, T. (2020). Public perceptions of good data management: Findings from a UK-based survey. *Big Data & Society*, 1–16. https://doi.org/10.1177/2053951720935616

Janssen, H., Cobbe, J., Norval, C., & Singh, J. (2020). Decentralised data processing: Personal data stores and the GDPR. *International Data Privacy Law*, *10*(4), 356–384. https://doi.org/10.2139/ssrn.3570895

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087

Mahieu, R., & Ausloos, J. (2020). *Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access* [Preprint]. LawArXiv. https://doi.org/10.31228/osf.io/b5dwm

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2), 205395172094808. https://doi.org/10.1177/2053951720948087

Mulgan, G., & Straub, V. (2019). *The new ecosystem of trust: How data trusts, collaboratives and cooperatives can help govern data for the maximum of public benefit* [Report]. Nesta. https://www.nesta.org.uk/blog/new-ecosystem-trust/

Ng, I., & Haddadi, H. (2018). Decentralised AI has the potential to upend the online economy. *WIRED*. https://www.wired.co.uk/article/decentralised-artificial-intelligence

O'Hara, K. (2019). Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship. *WSI White Paper*, *1*. https://doi.org/10.5258/SOTON/WSI-WP001

Pistor, K. (2020). Rule by data: The end of markets? *Law & Contemporary Problems*, *83*, 101–124.

Reed, C., B.P.E. Solicitors, & Pinsent Masons. (2019). *Data trusts: Legal and governance considerations*. Open Data Institute. http://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf

Ruhaak, A. (2019). Data trusts: Why, what and how. *Medium*. https://medium.com/@anoukruhaak/data-trusts-why-what-and-how-a8b53b53d34

Scassa, T. (2020). Designing data governance for data sharing. *Technology and Regulation*, 44-56 Pages. https://doi.org/10.26116/TECHREG.2020.005

Stalla-Bourdillon, S., Thuermer, G., Walker, J., Carmichael, L., & Simperl, E. (2020). Data protection by design: Building the foundations of trustworthy data sharing. *Data & Policy*, *2*, e4. https://doi.org/10.1017/dap.2020.1

Wong, J., & Henderson, T. (2020). Co-creating autonomy: Group data protection and individual self-determination within a data commons. *International Journal of Digital Curation*, *15*(1), 16. https://doi.org/10.2218/ijdc.v15i1.714

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5