Decentralized Data Infrastructures for the New Digital Economy

Thomas Hardjono
MIT Connection Science & Engineering
Email: hardjono@mit.edu
13 May 2021

Abstract

The TCP/IP Internet has provided the *communications infrastructure* for billions of people worldwide, with numerous social and economic benefits to humanity. However, as society becomes increasingly data-driven, we face a number of challenges regarding data ownership, centralization and privacy. A new paradigm for trusted *decentralized data infrastructures* is needed that balances the creation of value for data owners with the protection of user privacy [1].

The Data Driven Society

The various data regarding human behavior have always been essential for the functioning of government, industry, and society in general. Better insight is obtained when different types of data from various areas or verticals are combined. These insights allow communities to begin addressing the difficult social challenges of today, including better urban design, containing the spread of diseases, addressing economic inequality, and other challenges of the data-driven society [2].

However, in order for data for public goods to be readily available, the issues around personal data privacy and centralized control need to be addressed at the legal, social, and technical levels. The 2011 World Economic Forum (WEF) report on personal data [3] finds that the current ecosystems that access and use personal data are fragmented and inefficient:

- *Data is Siloed*: This makes data unavailable to support good decision-making.
- *Privacy is inadequately addressed*: For many participants, the risks and liabilities exceed the economic returns.
- *Security is failing*: The current Internet "firewalls" model is fundamentally inadequate, as evident by the almost daily reports of hacking and lost customer data.

To achieve an equitable data-driven society, we need a *new deal on data* – one with workable guarantees that the data needed for public goods are readily available while at the same time protecting the citizenry. This means, among others, to (i) treat personal data as an asset, (ii) to provide individuals with ownership rights in data that are about them regardless of what entity collects the data, and (iii) to provide individuals with full control over the use of their data, giving them the capacity to use data for their own purposes.

Data Cooperatives

One promising model that could be the foundation for implementing the new deal on data is the *data cooperative* [4], which is akin to the credit union organizations and farmers' co-ops in the United States. The notion of a data cooperative refers to the voluntary collaborative pooling by individuals of their personal data for the benefit of the membership of the group or community. The motivation for individuals to get together and pool their data is driven by the need to share common insights across data that would be otherwise siloed or inaccessible. These insights provide the cooperative members as a whole with a better understanding of their current economic, health and social conditions as compared to the other members of the cooperative generally. Some attractive features of the data cooperative include that it is member-owned, voluntary, and has legal *fiduciary obligations* to the individual members with regards to their data.

Decentralized Data Infrastructures: Goals

Just as DARPA's goals [5] for the TCP/IP Internet yielded some core design principles for the Internet architecture, today we are in need of coherent goals to motivate the exploration of new decentralized data infrastructures that can yield enduring design principles for the future data-driven society.

Possible goals could include some of the following [1]:

- *Support for limited data movement*: Support data minimization by storing/processing data *in situ* in its domain of origin. Algorithms should move to the data.
- *Support for distributed federated processing*: Support federated processing models that utilize distributed data repositories that protect privacy, where each may be owned by different entities and each may be protected using different technical means.
- *Support for distributed safe computations*: Employ distributed privacy-preserving computation techniques that protect data in storage, transmission and in computation.
- *Support for continuous audit*: Support continuous tracking and reporting of data usage and algorithm identification, yielding verifiable evidence to data-subjects.
- *Support for secure universal access*: Secure consent-driven authorized access based on robust digital identities must be supported, regardless of whether the endpoints are organizational data repositories or individual personal data stores (e.g. in data cooperatives). The advantages of digital infrastructures are diminished without universal access to data needed for public goods.

References

[1] A. Pentland, D. Shrier, T. Hardjono, and I. Wladawsky-Berger, "Towards an Internet of Trusted Data," in *Trusted Data - A New Framework for Identity and Data Sharing*, T. Hardjono, A. Pentland, and D. Shrier, Eds. MIT Press, 2019, pp. 15–40.

[2] D. Lazer et al., "Computational Social Science: Obstacles and Opportunities," *Science*, vol. 369, no. 6507, pp. 1060–1062, August 2020. [Online]. Available: https://doi.org/10.1126/science.aaz8170

[3] WEF, "Personal Data: The Emergence of a New Asset Class," World Economic Forum, Report, February 2011. [Online]. Available: http://www.weforum.org/reports/personal-data-emergence-new-asset-class

[4] A. Pentland and T. Hardjono, "Building Data Cooperatives," in *Building the New Economy: Data as Capital*, A. Pentland, A. Lipton, and T. Hardjono, Eds. MIT Press, 2021, pp. 19–33.

[5] D. Clark, "The Design Philosophy of the DARPA Internet Protocols," *ACM Computer Communication Review – Proc SIGCOMM 88*, vol. 18, no. 4, pp. 106–114, August 1988.