GLOSSARY ENTRY

OPEN ACCESS

PEER REVIEWED

Data cooperative

Alexander Fink *Augsburg University*

Abstract: Data cooperatives are a type of data intermediary using long established cooperative models and global regulatory affordances to manage data with or on behalf of members (data producers) for the benefit of members or for trade with external queriers (data users).

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

Introduction

Data, cheaply produced, stored, analysed, and sold, is a significant part of the financial basis upon which big tech platforms are constructed and by which they maintain significant power in the global economy (Srnicek, 2017). Centralisation places data and metadata of users in the hands of a few companies (Amazon, Apple, Alphabet, Facebook, and Microsoft) that buy, sell, link, and use it to manipulate, control, and advertise to users, while subjugating other businesses (Birch & Cochrane, 2022). This phenomenon is referred to as *platform capitalism* (Srnicek, 2017). Data cooperatives, a type of data intermediary emerging from cooperative organisational structures more generally(Scholz & Schneider, 2016), represent a user- or worker-owned alternative that suggests that data, and the economic and social gains emerging from it, be owned by users or workers — a decentralised form of data ownership that promotes collective power alongside individual or group owner-autonomy over data (Miller, 2021). In the specific form of the *platform cooperative*, data cooperatives occupy analogous positions as under platform capitalism, while changing ownership structures and foregrounding ethical data use (Scholz, 2023). Beyond data privacy and user control over data use in areas like health and science, other forms and purposes for data cooperatives exist, including the management of data across small businesses or sole proprietorships, in areas like farming (e.g. SAOS and GISC), or potentially the non-profit sector. Definitions and examples of data cooperatives are evolving rapidly through experimentation and regulatory changes, and there are a diverse array of entities referring to themselves as *data cooperatives*, when in fact some may more accurately be another kind of data intermediary (European Commission et al., 2023, identify 6 types of data intermediaries[1]) because they are not managed through cooperative enterprises, which are a specific legal entity in most countries globally (Gordon Nembhard, 2014).

Cooperative ownership of enterprises is a long-standing tradition around the world and data cooperatives derive standard practices and governance models from this history (Cheney et al., 2023), while also seeking tools, information system architectures, and governance models specific to data capture/creation, management,

---

1. The six types identified are: (1) Personal Information Management, Data Cooperatives, Data Trusts, Data Unions, Data Marketplaces, and Data Sharing Pools (European Commission et al., 2023).

analysis, and use (Pentland et al., 2019). Data cooperatives lean on state regulatory affordances and economic support for cooperative organisations, as well as cooperative governance techniques used in organisations for dozens of years around the world (Micheli et al., 2020). Nevertheless, research on data cooperatives is new and a limited number of data cooperatives exist.

Definition and origin of the term

In the current digital environment, the vast majority of data is created, harvested, and/or traded by private corporations and governments, and used and traded on their behalf. In contrast, open data promotes the public release of data to all parties for any purpose, including the private benefit of corporations. In parallel with a long history of cooperatively owned and governed enterprises supported by regulatory systems around the world, data cooperatives provide a technical, legal, and governance structure for the collective management and distribution of data on behalf of cooperative members. Cooperatives differ also from *data commons*, in the sense that the latter does not identify with the international cooperative movement or corresponding legal affordances offered in many countries.

Some data cooperatives may emulate cooperative enterprises without benefiting from the legal affordances available in most countries. Many, however, follow the International Cooperative Alliance (ICA) approach, which defines a cooperative as "an autonomous association of persons united voluntarily to meet their common economic, social and cultural needs and aspirations through a jointly-owned and democratically controlled enterprise" (ICA, n.d.). The ICA suggests seven core principles held in common globally in the cooperative movement, though others have suggested modifying or adding to these (Cheney et al., 2023; distributed cooperative organisation, 2019). These are: (1) voluntary membership, (2) democratic member control, (3) economic participation of members, (4) autonomy and independence, (5) education, training and information, (6) cooperation among cooperatives, and (7) concern for community (Cheney et al., 2023, pp. 23–24). Cooperatives are usefully differentiated by ownership group, including worker-, producer-, and consumer-owned cooperatives (Cheney et al., 2023). Cooperative managers or boards have a fiduciary responsibility to members, including duties to manage (achieve the work of the coop), care (to do what is best for the coop), loyalty (first allegiance is to the cooperative), and obedience (comply with state regulations). Legally, cooperatives are regulated by state policy in most countries around the world (Gordon Nembhard, 2014).

Data cooperatives are a specific type of cooperative, acting as an intermediary or-

ganisation (also known as a *data intermediary*) collecting data from members and trading it to accrue value *for members*. Data may be of many types, ranging from personal — like browsing and social media, health and fitness (Tanwar et al., 2021), shopping, or banking — to environmental, including construction management (Bühler et al., 2023), defence against phishing (Salau et al., 2021), or citizen science projects (Tanwar et al., 2021). Existing data cooperatives tend to be producer-owned, in the sense that those providing data and benefiting from its trade are the "producers". Membership in data cooperatives for individuals may require contributing data and/or other resources, like a monetary fee (Cheney et al., 2023). Other data cooperatives may be cooperatives of organisations, wherein companies or non-profits are members of a broader data cooperative, as is the case in a proposed construction management cooperative (Bühler et al., 2023).

Once held cooperatively, data may be used, managed, and shared, serving many purposes. The largest and most successful data cooperatives to date may be health data cooperatives, where individuals upload and share health data, for the purposes of informing their doctors and supporting scientific research and intervention development, while maintaining granular control over their personal data (retaining the ability to limit or revoke access). Thus, rather than consent being given at the point of collection (if at all) and data disappearing into databases that are inaccessible to regular users, systems like MIDATA allow users to change access permissions at any time and to understand which groups and for what purpose their data are being used (Hafen et al., 2014; Tanwar et al., 2021). Users can therefore be more actively involved in projects or apps that use their data, and third party services are incentivised to take consent, communication, and involvement with users more seriously. Health data cooperatives may have the potential to provide additional benefits to doctors and researchers by reducing some barriers to data sharing, although this possibility is not yet verified. Such cooperatives may also increase cooperative members' understanding of where and how their data is to be used (Hafen et al., 2014).

Individuals join data cooperatives to gain greater power and control over data sharing, as well as for data altruism — the ability to collectively and consensually share personal data toward altruistic causes, like pharmaceutical research (Bietti et al., 2021). Some cooperatives may be basic and require little participation from members, while others may involve significant participation in individual data management and governance (Hafen et al., 2014; Hardjono & Pentland, 2019; Micheli et al., 2020).

Data cooperatives share common policy regulation, governance and membership,

and core principles with cooperatives generally (Cheney et al., 2023; Gordon Nembhard, 2014). However, the specific nature of data as the cooperatives' product, and the distributed and digital nature of organisation, present unique challenges and opportunities for data cooperatives, such as granular management of governance and business processes (like voting on management decisions or changing consent to do business with certain entities), foregrounding the importance of security in operations, and distinct regulatory frameworks (Hardjono & Pentland, 2019; European Commission et al., 2023; Scholz & Calzada, 2021).

Component concepts

To be cooperatively held and managed, data cooperatives must establish technical systems as well as sociocultural governance practices (Ruijer, 2021). Technical systems include cloud storage, computing infrastructure, consent management, front and back end applications to access, manipulate, and analyse data, and security considerations. Governance and policy refer to internal governance, as well as regulatory policies for data intermediaries, data, and cooperatives.

Socio-technical considerations

Technical systems vary widely (Bühler et al., 2023; distributed cooperative organisation, 2019; Hardjono & Pentland, 2019) and must address specific considerations around at least four issues: sharing management (Bühler et al., 2023; Hafen et al., 2014), consent systems, security (Salau et al., 2022), and state policy requirements related to data protection and data intermediaries (e.g. European Union through Data Governance Act) (Bietti et al., 2021; European Commission et al., 2023). Such systems are inherently also social and organisational, because technical affordances reciprocally determine user interactions.

Each of these areas represent challenges for data cooperatives around which there are not yet general agreements, though some have proposed socio-technical solutions. Regarding the management of data sharing, most cooperative systems generally organise around three basic user roles:

- Personal data storers / cooperative members: members store personal data and make decisions about its use and sharing, including giving legal consent where necessary (e.g. healthcare data privacy regulations), managing sharing at variously granular levels, adding/modifying/ destroying data or data streams in a personal data store, electing cooperative leadership or participating in management decisions, and receiving benefits from the cooperative (e.g. financial rewards). Data stores

    are typically managed by the cooperative body on behalf of members, but decentralised technologies may also be employed.

- Queriers: users trading for access to data owned by the cooperative. Access includes the ability to run queries against the data (hence their role title), run algorithmic analysis on cooperative data, and pay for services rendered.
- Data cooperative: data cooperative systems must manage user needs for cooperative members and queriers, providing security for both parties in data storage, use, and financial trade, systems for managing data storage, and robust and useful systems for making queries (Blasimme et al., 2018; Hardjono & Pentland, 2019).

Consent systems promote user privacy, security, and control over data about them. These often intersect with regulatory requirements, including the General Data Protection Regulation (GDPR) in the European Union, which governs data generally, and Health Insurance Portability and Accountability Act (HIPAA) in the United States, which manages health care data. Developing systems that allow members more granular control over data sharing at any time increases member consent (Fink & Brito, 2021), while simultaneously decreasing queriers barriers to combining data across multiple institutions, which otherwise might require separate consent agreements across each institution (Hafen et al., 2014). Queriers algorithms must be vetted to ensure member safety and consent, as well as to prevent bias in algorithms (Hardjono & Pentland, 2019).

Individual members should be able to revoke access at any time, requiring that data remain in control of the cooperative at all times. The MIT Open Algorithms approach solves this problem by requiring queriers to submit algorithms and queries to the cooperative to be run, preventing queriers from taking, owning, or linking cooperative data to others (Bietti et al., 2021; Hardjono & Pentland, 2019).

Security in data cooperatives is under-analysed but significant. Developing large, intensely personal data stores, alongside storing querier algorithms, sets up data cooperatives as a target for hackers. Spoofing, tampering, repudiation, information disclosure, denial-of-service, and elevation of privilege attacks must all be considered in developing a threat analysis and security plan (Salau et al., 2022).

Policy and regulatory considerations

Policies and regulations impact the operation of cooperatives, as well as the storage, management, use, and sharing of data. Globally, few protections exist for individuals to protect privacy or abuse of information by corporations. The European Union's GDPR act is one of the stronger (inter)national policies. The European

Union's recent Data Governance Act "aims at increasing trust in voluntary data sharing" by regulating data intermediaries, including limitations to reselling data for profits (a common practice of platform capitalism) (European Commission et al., 2023, p. 7). The United Kingdom and United States have industry specific data protection laws (Bietti et al., 2021), but lack the more serious protections and affordances of the European Union. The African Union recently published a Data Policy Framework, which suggests some support for privacy protections (African Union, 2022). A global web of national privacy regulations creates significant difficulties in setting up international cooperatives, limiting most to national operations. Some propose federated data cooperatives (a cooperative of cooperatives), aggregating members within the boundaries of state regulations, and then sharing data between state bodies (Tanwar et al., 2021).

Data cooperatives are thus far under-resourced and researched, partly due to challenges creating sustainable economic models, developing user trust in data sharing, and little to no financial support from governments (European Commission et al., 2023). Policy could promote data cooperatives as a vehicle to support public empowerment and involvement in data use, data privacy, economic development, and sharing of important data resources. The best existing example is the Data Governance Act (DGA), which regulates data intermediaries in the European Union to build transparency in intermediary operations, but also provides affordances, like public registers of intermediaries to increase visibility and protections against misuse, in hopes these will increase public trust in data sharing (European Commission et al., 2023). Elsewhere, to date, little government support exists for the foundation or growth of data cooperatives (Bietti et al., 2021).

Economic considerations

Thus far, little information exists on the economic viability of data cooperatives. Monetisation models differ by sector. Some require members to join related cooperatives or pay membership fees/dues (e.g. MIDATA, SAOS, GISC). These organisations may rely entirely on membership dues (e.g. SAOS) or seek other sources of funding. Some data cooperatives receive startup funds from foundations (e.g. Posmo) or corporations. These may intend long-term operations based in foundation or corporate funding, while others aim to develop viable financial models based on the sale of data to external users for research, corporate, or other purposes (e.g. MIDATA). This model faces significant challenges, especially because meaningfully large datasets are difficult to create without successfully incentivising participation. While platform capitalist models spend significant financial investments to create data streams that often engage millions of users, data cooperatives must ei-

ther find niche markets (e.g., GISC, which serves food growers) or try to build momentum through multi sectoral partnerships and marketing to gain a critical mass (e.g. MIDATA).

Data cooperatives are often related to approaches to Social and Solidarity Economics (SSE), which the United Nations recognises as encompassing "entities that are engaged in economic, social and environmental activities to serve the collective and/or general interest, which are based on the principles of voluntary cooperation and mutual aid, democratic and/or participatory governance, autonomy and independence and the primacy of people and social purpose over capital in the distribution and use of surpluses" (UN Resolution 77/281, p. 2). The democratic nature of governance and ownership, voluntary participation and consent of members and general focus on aiding public interest align data cooperatives closely with SSE aims. This puts data cooperatives in line with (and against) national and global political and policy interests, wherein such cooperatives may be able to find amenable funding and government support.

To ease the challenge of establishing new organisations for the management of data cooperatives, Pentland et al. (2019) suggest utilising existing organisations with similar structures and fiduciary responsibilities and the capabilities to manage technical challenges present in data management specifically. The credit union is suggested as the most readily capable organisation, with existing fiduciary responsibilities to manage duties and care for members, existing memberships and member management systems, security, and near universal legal frameworks globally (Pentland et al., 2019). Others, including health data cooperatives, create new entities using state regulations around cooperative ownership (e.g. SalusCoop, MIDATA)(Blasimme et al., 2018).

Governance considerations

Many robust models for member governance have been developed, tested, and shared (Cheney et al., 2023; Gordon Nembhard, 2014). Governance and member participation can, but does not always, occur along multiple axes, including decision-making at all levels (strategic direction to technical implementation), involvement in debating and solving key issues, and/or in choosing cooperative management (Cheney et al., 2023). These member governance activities may intersect with the technical roles presented above, as, for example, data cooperatives often have technical systems for managing member voting. Many producer-owned cooperatives elect board members or hire employees to manage the regular tasks of the cooperative. Regardless, education in cooperative principles and practices, es-

pecially in democratic practice, is arguably one of the most important tasks for long-term sustainability (Cheney et al., 2023).

These governance practices are true for data cooperatives too, where the distributed and digital nature of cooperative membership can present special challenges and opportunities. While many data cooperatives reduce member work burdens by voting for board representation and management, others seek to use blockchain tools or other technological affordances to govern and operate the cooperative using methods more akin to direct democracy (distributed cooperative organisation, 2019; Ruijer, 2021).

Discussion and conclusion

As research, policy, and practice around data cooperatives continue to develop, three areas of recent significance in the broader cooperative movement may prove valuable for consideration: inclusivity, education and ecological/environmental (Cheney et al., 2023). Historically, people around the world and across many identity categories have been part of the cooperative movement (Gordon Nembhard, 2014). However, efforts to develop data cooperatives that attend to issues of inclusion and equity are limited. Cheney et al. (2023) assert that education in cooperative governance and methods is essential to developing and maintaining cooperatives as such in the long-term; otherwise, many cooperatives devolve into standard, hierarchically governed organisations. Ecological considerations were not historically part of the cooperative movement, but are being added by cooperatives and even data cooperatives for consideration (Cheney et al., 2023; distributed cooperative organisation, 2019). These areas are only tangentially addressed thus far in the literature on data cooperatives, but meaningful investigation could assist the further development of the area.

Although data cooperatives offer more democratic models of data sharing and governance, there are several limitations at present. The relative newness of data cooperatives leaves open questions about the long-term sustainability and security of data in the event that a cooperative must fold. Governance and sharing models are relatively untested and little research exists on their effectiveness and inclusivity. Many data cooperatives depend on broad membership or participation in data sharing, and it remains unclear whether, for what reasons, and in which contexts broader publics will be motivated to share data. Finally, data cooperatives in many industries will face significant competition from corporate firms with more significant capital.

Despite these challenges, data cooperatives remain a worthy target for further development and research. Data cooperatives have the potential to provide a platform for member-owners to create mechanisms that simultaneously enhance privacy and control over their digital exhaust, while also enabling desirable access for profit or altruistic goals. In theory, they provide easy, consistent access for queriers to access data in ways that are safe, secure, and built on owner consent and active participation, though socio-technical systems enabling such queries are relatively new. The political and social medium for these enterprises exists globally and presents significant opportunities in many industries such as health research, participatory governance, and small business alliances, as noted above. Other industries may benefit from joining workers or data producers together as cooperatives, including the non-profit and human services sectors, where cooperative systems could help manage services and funders (Fink & VeLure Roholt, 2022).

References

African Union. (2022). *AU data policy framework* [Report]. https://au.int/en/documents/20220728/au-data-policy-framework

Bietti, E., Etxeberria, A., Mannan, M., & Wong, J. (2021). *Data cooperatives in Europe: A legal and empirical investigation* (The New School's Platform Cooperativism Consortium) [White paper]. https://cyber.harvard.edu/sites/default/files/2022-02/Data_Cooperatives_Europe-group2.pdf

Birch, K., & Cochrane, D. T. (2022). Big tech: Four emerging forms of digital rentiership. *Science as Culture*, *31*(1), 44–58. https://doi.org/10.1080/09505431.2021.1932794

Blasimme, A., Vayena, E., & Hafen, E. (2018). Democratizing health research through data cooperatives. *Philosophy & Technology*, *31*, 473–479. https://doi.org/10.1007/s13347-018-0320-8

Bühler, M. M., Nübel, K., Jelinek, T., Riechert, D., Bauer, T., Schmid, T., & Schneider, M. (2023). Data cooperatives as a catalyst for collaboration, data sharing and the digital transformation of the construction sector. *Buildings*, *13*(2), Article 442. https://doi.org/10.3390/buildings13020442

Cheney, G., Noyes, M., Do, E., Vieta, M., Azkarraga, J., & Michel, C. (2023). *Cooperatives at work.* Emerald Publishing Limited. https://doi.org/10.1108/9781838678258

distributed cooperative organisation. (2019). *If I only had a heart: A DisCO manifesto* [Manifesto]. https://manifesto.disco.coop/

European Commission, Joint Research Centre, Micheli, M., Farell, E., Carballa-Smichowski, B., Posada-Sánchez, M., Signorelli, S., & Vespe, M. (2023). *Mapping the landscape of data intermediaries: Emerging models for more inclusive data governance.* [Report]. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/8943

Fink, A., & Brito, M. (2021). Real big data: How we know who we know in youth work, child & youth services. *Child & Youth Services*, *42*(2), 150–178. https://doi.org/10.1080/0145935X.2020.1832888

Fink, A., & VeLure Roholt, R. (2022). Emerging tensions in data work: Staff and youth perspectives in youth-serving organizations. *Social Service Review*, *96*(4), 617–654. https://doi.org/10.1086/7222 77

Gordon Nembhard, J. (2014). *Collective courage: A history of African American cooperative economic thought and practice*. Penn State University Press. https://doi.org/10.5325/j.ctv14gpc5r

Hafen, E., Kossmann, D., & Brand, A. (2014). Health data cooperatives—Citizen empowerment. *Methods of Information in Medicine*, *53*(02), 82–86. https://doi.org/10.3414/ME13-02-0051

Hardjono, T., & Pentland, A. (2019). *Data cooperatives: Towards a foundation for decentralized personal data management* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1905.08819

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), 1–15. https://doi.org/10.1177/2053951720948087

Miller, K. (2021, October 20). Radical proposal: Data cooperatives could give us more power over our data. *Stanford University Human-Centered Artificial Intelligence*. https://hai.stanford.edu/news/rad ical-proposal-data-cooperatives-could-give-us-more-power-over-our-data

Pentland, A., Hardjono, T., Penn, J., Coclough, C., Ducharme, B., & Mandel, L. (2019). *Data cooperatives: Digital empowerment of citizens and workers* [White paper]. MIT Connection Science. htt ps://ide.mit.edu/sites/default/files/publications/Data-Cooperatives-final.pdf

Ruijer, E. (2021). Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly*, *38*(4), Article 101612. https://doi.org/10.1016/j.giq.2021.101612

Salau, A., Dantu, R., Morozov, K., Upadhyay, K., & Badruddoja, S. (2022). Towards a threat model and security analysis for data cooperatives: *Proceedings of the 19th International Conference on Security and Cryptography*, 707–713. https://doi.org/10.5220/0011328700003283

Salau, A., Dantu, R., & Upadhyay, K. (2021). *Data cooperatives for neighborhood watch*. 1–9. https://d oi.org/10.1109/ICBC51069.2021.9461056

Scholz, R. T. (2023). *Own this!: How platform cooperatives help workers build a democratic internet*. Verso.

Scholz, T., & Calzada, I. (2021). Data cooperatives for pandemic times. *Public Seminar*. https://doi.or g/10.13140/RG.2.2.12320.51200/1

Scholz, T., & Schneider, N. (Eds.). (2016). *Ours to hack and to own: The rise of platform cooperativism, a new vision for the future of work and a fairer internet*. OR Books. https://doi.org/10.2307/j.ctv62hfq7

Srnicek, N. (2017). *Platform capitalism*. Polity Press.

Tanwar, A. S., Evangelatos, N., Venne, J., Ogilvie, L. A., Satyamoorthy, K., & Brand, A. (2021). Global open health data cooperatives cloud in an era of COVID-19 and planetary health. *OMICS: A Journal of Integrative Biology*, *25*(3), 169–175. ht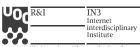tps://doi.org/10.1089/omi.2020.0134