12-2022

Toward Cross-company Value Generation from Data: Investigating the Role of Data Sharing Communities

Hippolyte Lefebvre

Pavel Krasikov

Gabin Flourac

Christine Legner

Toward Cross-company Value Generation from Data: Investigating the Role of Data Sharing Communities

Short paper

Hippolyte Lefebvre
Faculty of Business and Economics (HEC), University of Lausanne
1015 Lausanne, Switzerland
hippolyte.lefebvre@unil.ch

Pavel Krasikov
Faculty of Business and Economics (HEC), University of Lausanne
1015 Lausanne, Switzerland
pavel.krasikov@unil.ch

Gabin Flourac
Faculty of Business and Economics (HEC), University of Lausanne
1015 Lausanne, Switzerland
gabrin.flourac@unil.ch

Christine Legner
Faculty of Business and Economics (HEC), University of Lausanne
1015 Lausanne, Switzerland
christine.legner@unil.ch

Motivation and Research Objective

Without a doubt, data is considered as a strategic asset for the digital economy. While companies seek for greater data-driven insights to unlock new business opportunities, we observe a shift from internal to cross-company data sharing (Wixom, Sebastian and Gregory, 2020). The European data strategy (2020), and the underlying EU's data act (2022) bring upfront numerous benefits of data sharing, such as improved access to private and public data, generation of new products and services, or reduction of public services' costs, amounting to 270 billion euros in additional GDP by 2028. In fact, estimates from Gartner (2021) show that private organizations engaging in sharing their data can expect to generate three times more measurable economic benefit compared to those who do not. Data sharing also contributes to the sustainable use and reuse of data particularly in the context of reduction of energy use and technological resources (European Commission, 2020; Akhgarnush, 2021)

In recent years, novel forms of data sharing have emerged where firms share data horizontally with others (sometimes competitors), as opposed to the more traditional concept of data exchange which creates benefits by sharing data vertically with all the actors of the value chain, for instance, a supplier can exchange data with retailers, distributors, or other sales intermediaries to derive data driven decisions or insights (Otto *et al.*, 2019). Companies might then share their data horizontally in a community of peers, sharing common practices and seeking for mutual benefits, such as cost-sharing or operational efficiency (Wixom, Sebastian and Gregory, 2020). These data sharing communities can be perceived as closed ecosystems and referring to the concept of gated communities and club goods which have strong boundaries, a restricted access (Blakely, 2006), and where the outcome(s) of the community must benefit only its members (Buchanan, 1965; Nicolini *et al.*, 2022). Despite the relevance of data sharing, research still lacks a thorough conceptualization of data sharing communities. Since practice exchange around a shared domain of interest is pivotal for data sharing, we suggest to investigate it using communities of practices (CoP) as the theoretical lens (Wenger, 1998). Hence, we formulate the following research question: *How do communities of practice support data sharing?*

Pursuing our inquiry, we use a multiple case study research design, which is appropriate when few research studies have looked into the topic of interest (Benbasat, Goldstein, and Mead, 1987). Our research allows us to propose an academic conceptualization of data sharing communities built upon the well-established

theoretical lens of CoPs. We find that companies might combine vertical and horizontal data sharing to augment their value chain, or focus only on horizontal data sharing to improve value of data through data management activities. Hence, we contribute to the emerging body of knowledge on cross-enterprise data sharing by showing how data sharing communities are relevant for companies to mutually benefit from data. For practitioners, we propose empirical evidence on how companies are built and operated.

Background

The concept of data sharing is primarily associated with providing a facilitated access for use and reuse of data to every stakeholder at any time (Tenopir *et al.*, 2011). To this day, there is no consensus on data sharing's academic definition (Graef and Prüfer, 2021). One of the reasons for this, as explained by Wixom et al. (2020), is the evolution from conservative practices of sharing minimal data (data sharing 1.0) to cross-company sharing of complementary data assets (data sharing 2.0). The first one is primarily driven by enterprises' needs to preserve an existing value proposition (e.g., execute a transaction or comply with a regulation), while the second aims at an improved value creation (Wixom, Sebastian and Gregory, 2020). In the context of data ecosystems, data sharing's main trait is that it primarily takes place in "horizontal cooperation and collaboration between companies", rather than vertically, i.e. in data exchanges (Otto *et al.*, 2019). European Commission's data act, made companies become more aware that data needs to be shared outside the organizations' boundaries in order to avoid silos, and unearth new economic and social benefits. However, the emergence of cross-company data sharing paradigm is confronted with new challenges and barriers, for instance, public and private actors being reluctant to share data outside their organization due to a misalignment of incentives, privacy concerns, and a lack of collaboration between members (Skatova, Ng and Goulding, 2014; Taylor and Mandl, 2015; Klein and Verhulst, 2017; Liu *et al.*, 2017; Susha, Janssen and Verhulst, 2017; van den Homberg, 2017). While the economic benefits have been primarily discussed on the strategic level (European Commission, 2020), literature mentions only few benefits from the enterprise perspective, e.g., the minimization of costs related to the data re-collection (Tenopir *et al.*, 2011), leading to economies of scale. For instance, members may be competitors but share their data to optimize processes, derive insights and extract value out of data at the same level of the value chain (Otto et al., 2019), which extends the sharing economy principles to data (Legner, 2019).

Overall, research still lacks a proper conceptualization of data sharing communities in order to outline their role in modern data sharing. Cross-organization data sharing still remains misunderstood in literature in terms of organizational norms and terminology, oftentimes being referred to as data collaborative, data donation, data partnership, or even data exchange (Hale *et al.*, 2003; Susha, Grönlund and Van Tulder, 2019). Besides, while actors in general data sharing– data consumer, data provider and service provider - are clearly identified in literature (Schlosser, 2016), they mainly appear in the context of unilateral data exchange. This does not account for the need to observe data sharing as a bidirectional flow between provider and consumers, endorsing the interchangeable role of *data prosumer* (Otto and Aier, 2013). In such setting, data prosumers and service providers ensure facilitation (Susha, Janssen and Verhulst, 2017) and form a closed data and metadata-sharing ecosystem, built upon a data pool with community members.

Methodology

We opted for analyzing multiple exploratory case studies (Yin, 2003) to improve our understanding of data sharing between organizations. Multiple case studies are particularly relevant to derive a qualitative understanding about a phenomenon of interest in its natural setting and to capture as much knowledge as possible from practitioners (Paré, 2004). Furthermore, multiple case studies are appropriate when few research has been produced for the topic of interest (Benbasat, Goldstein, and Mead, 1987; Yin, 2003). In the following sections, we elaborate on our research methodology.

Case Selection and Data Collection

To identify data sharing cases, we searched for prominent and mature data sharing communities which would be diverse enough in terms of their industry focus, actors, and size. We first browsed and researched examples of data sharing communities on the website of Support Centre for Data Sharing (2022), the European Commission's initiative to facilitate the knowledge exchange and provide data sharing practice examples. By browsing through the list of use cases, we could also filter our search per industry sector (e.g.,

mobility, agriculture). Besides using the website's search engine, we used the tags "Data sharing" and "Data sharing community" to search for most relevant examples, resulting in 20 candidate communities. After exploring the latter list, we filtered out communities that had been recently set up i.e., created after 2020, were not fully operational or in a pilot stage. Furthermore, in line with purposeful sampling (Patton, 1990), we only selected "information-rich" communities. We also ensured to have diverse cases in the sample in terms of industries, years of registration and data sharing models. This resulted in a sample of five data sharing communities : 1) *Business Partner Data Sharing Community* operated by CDQ AG since 2017; 2) *Skywise* operated by Airbus since 2017; 3) *Smart Farming* operated by Join data since 2017; 4) *MAAS Madrid* (recently rebranded as EMT Mobility 360) operated by EMT Madrid since 2018; and 5) *Benchmarking for Industry Association* operated by Data Sharing Coalition since 2020 (additional information in Table 1). Once the sample identified, we gathered information about them directly from the Support Center for Data Sharing website or from the dedicated websites of each initiative.

Within and Cross-Case Analysis

We first analyzed the individual cases by coding them against our theoretical framework, built upon the CoP lens (Wenger, 1998). After the within-case analysis, we analyzed commonalities and differences between the cases to identify the patterns.

Preliminary Results

Our findings show that data sharing is typically performed in communities of peers composed of data experts from respective companies, who share a common concern or a passion and learn how to improve by exchanging the shared practices. Each community member is at the same time consumer and provider of data. Therefore, data sharing communities can be characterized along the three structural elements of CoP (Wenger et al., 2002): Community, Shared Practices and Domain of Interest. This allows us to propose a first academic definition for data sharing communities: we define a data sharing community as a community composed of selected organizations (community members) sharing the same domain of interest who interact, collaborate, and commonly share/use any type of data to achieve a common goal and benefit from the created added value.

Table 1 synthesizes our analysis of the five cases, i.e., how each case has been assessed against this theoretical framework. We further broke down Wenger's characteristics of CoP along the relevant sub-characteristics to investigate data sharing (Susha, Janssen and Verhulst, 2017): the *shared data types*; the *facilitation* i.e., indicating potential existing intermediary and its role in the community; *access to shared data*. Interestingly, despite a common general goal of sharing data between different parties, we notice significant alterations between them and can distinguish two patterns. First, we notice that the domain of interest (or goal) for most communities (Skywise, Smart Farming, Maas Madrid, Benchmarking for Industry Association) is about supporting direct or indirect value creation from the shared data. For example, the data is shared by providers, so that consumers can generate analytics insights or business insights. Skywise and Airbus, with their suppliers and customers (e.g., airlines), continuously track and analyze operations and performance data from aircrafts. This can be considered an extension of vertical data sharing, where the focus is about joint data use and innovation. Conversely, for the CDQ Business Partner Data Sharing Community, actors share only master data (e.g., from common suppliers), so that data creation and maintenance costs are split, increasing the efficiency. Thereby, all members provide their data into a shared data pool, followed by assessment and enrichment for data quality improvement. In this case, members contribute to the elaboration of a shared practice, for instance by developing new data quality rules. Since member are from different industries and are not part of the same value chain, this is an example of purely horizontal data sharing. We notice that in all analyzed cases, the community development and operations are facilitated by an intermediary who provides organizational and technical support.

CoP characteristics		Business Partner Data Sharing Community	Skywise	Smart Farming	Maas Madrid	Benchmarking for Industry Association
Domain of interest	**Goal**	Maintain and enrich business partner master data for cost-sharing	Derive valuable insights for improving plane components or anticipate maintenance	Agricultural innovation. Derive data driven insights, improve processes and profitability and optimize their production	Optimize and improve the public transports efficiency	Provide members and the larger public with aggregated insights and trends in industry performance
	Industry focus	Cross-industry	Airlines and aircrafts manufacturing	Farming and associated research	Public transportation	Various companies with industry-specific relationship
Community	**Members (Consumers and providers)**	20 multinational companies from various industries	300 actors, both airlines having Airbus aircrafts and manufacturers/suppliers	10 actors from the agricultural sector e.g., farmers, laboratories, knowledge institutes and agricultural entrepreneurs.	70 mobility actors in Madrid who share their data (e.g., IoT, Geo)	60 actors from different related industries
	Nature of data sharing	Private-Private	Private-Private	Private-Private	Public-Private	Private-Private
	Type of facilitation	Intermediary with data-related & organizational functions	Intermediary with data-related & organizational functions	Intermediary with data-related functions	Intermediary with data-related & organizational functions	Intermediary with data-related & organizational functions
Shared practices	**Practice focus**	Data management e.g., definition and standards, Data quality rules	Continuous tracking and analysis of operations and performance data	Cooperative-level analytics and secure exchange of agricultural data	Central repository for live mobility data allowing real time route optimization	Industry-specific data sharing use case design (e.g., data sharing interaction model, business and legal, requirements)
	Shared data types	Business partner master data, Financial data	Product master data, IoT data, Geo data	IoT data, Geo data	Product master data, IoT data, Geo data	Financial data
	Type of data sharing	Actors are prosumers	Actors are either providers or consumers	Actors are either providers or consumers	Actors are either providers or consumers	Actors are either providers or consumers
	Access to shared data	Access to outcomes of processed data, Access to modified or enrich data	Access to raw data, Access to outcomes of processed data	Access to outcomes of processed data, Access to modified or enrich data	Access to outcomes of processed data	Access to outcomes of processed data

Table 1. Mapping of cases against theoretical framework

Next Steps and Expected Contribution

As data sharing communities gain momentum, we believe that our preliminary results lay the foundation for further empirical analysis. They help to assemble the scattered body of knowledge focusing on specific

domains (e.g., industry) and casting various terminologies such as data collaborative (Susha, Janssen and Verhulst, 2017), data donation (Skatova, Ng and Goulding, 2014), data philanthropy (Taddeo, 2017) or data-driven social partnership (Susha, Grönlund and Van Tulder, 2019). We believe that clarifying terms and concepts is essential at a time when data sharing becomes increasingly important. Furthermore, the patterns that emerge from our cross-case analysis offer interesting opportunities for future research. On the one hand, most communities focus on supporting insights generation, hence providing further learnings to support the existing value chains. Furthermore, these communities leverage horizontal data sharing to augment an already in-place vertical data sharing. On the other hand, we find that data sharing also supports the increasing value of data as a shared asset, for instance by improving its quality. In this regard, the CDQ Business Partner Data Sharing Community proves to be of particular interest, as it is also the only community in our sample that focuses on the data itself (e.g., master data) and not the outcome of data processing (e.g., analytics). Accordingly, it provides opportunities for further understanding of horizontal data sharing in a longitudinal study of this community, and its development and operation. As one of the oldest communities in our case base, this will allow us to get an understanding of its success factors, development over time, and successful operation, coupled with maintaining sustained interest.

References

Akhgarnush, E. (2021) 'Data Sustainability—A Thorough Consideration', in V. Liermann and C. Stegmann (eds) *The Digital Journey of Banking and Insurance*. Cham: Springer International Publishing, pp. 119–129.

Benbasat, I., Goldstein, D.K. and Mead, M. (1987) 'The Case Research Strategy in Studies of Information Systems', *MIS Quarterly*, 11(3), pp. 369–386.

Buchanan, J.M. (1965) 'An Economic Theory of Clubs', *Economica*, 32(125), pp. 1–14.

European Commission (2020) *A European Strategy for data*. Available at: https://digital-strategy.ec.europa.eu/en/policies/strategy-data (Accessed: 1 October 2022).

European Commission (2022) *Data Act: Measures for a Fair and Innovative Data Economy*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113 (Accessed: 1 October 2022).

Gartner (2021) 'Data Sharing Is a Business Necessity to Accelerate Digital Business', 20 May. Available at: https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business (Accessed: 5 January 2022).

Graef, I. and Prüfer, J. (2021) 'Governance of Data Sharing: A Law & Economics Proposal', *Research Policy*, 50(9). Available at: https://linkinghub.elsevier.com/retrieve/pii/S004873332100130X (Accessed: 12 October 2021).

Hale, S.S. *et al.* (2003) 'Managing troubled data: coastal data partnerships smooth data integration', in *Coastal Monitoring through Partnerships*. Springer, pp. 133–148.

van den Homberg, M. (2017) 'Toward a Balkans' Data for Disaster Management Collaborative?', in *Implications of climate Change and disasters on Military Activities*. Springer, pp. 11–18.

Klein, T. and Verhulst, S. (2017) *Access to New Data Sources for Statistics: Business Models and Incentives for the Corporate Sector*. Paris: OECD. Available at: https://www.oecd-ilibrary.org/economics/access-to-new-data-sources-for-statistics_9a1fa77f-en (Accessed: 1 October 2022).

Legner, C. (2019) 'Data Sharing in Business Ecosystems', in *TREO Talks in conjunction with the Fortieth International Conference on Information Systems (ICIS)*. Munich, Germany.

Liu, Y. *et al.* (2017) 'Donating Context Data to Science: The Effects of Social Signals and Perceptions on Action-taking', *Interacting with Computers*, 29(2), pp. 132–146.

Nicolini, D. *et al.* (2022) 'Understanding Communities of Practice: Taking Stock and Moving Forward', *Academy of Management Annals*, 16(2), pp. 680–718.

Otto, B. *et al.* (2019) *Data Ecosystems. Conceptual Foundations, Constituents and Recommendations for Action*.

Otto, B. and Aier, S. (2013) 'Business Models in the Data Economy: A Case Study from the Business Partner Data Domain', in *Proceedings of the Wirtschaftsinformatik 2013*.

Paré, G. (2004) 'Investigating Information Systems with Positivist Case Research', *Communications of the Association for Information Systems*, 13(1), p. Paper 18.

Patton, M.Q. (1990) *Qualitative evaluation and research methods*. SAGE Publications, inc.

Schlosser, S.R. (2016) *Design principles for collaborative data services*. University of St. Gallen.

Skatova, A., Ng, E. and Goulding, J. (2014) 'Data Donation: Sharing Personal Data for Public Good', *Application of Digital Innovation. London, England: N-Lab* [Preprint].

Support Centre for Data Sharing (2022) *What is data sharing?* Available at: https://eudatasharing.eu/what-data-sharing (Accessed: 3 October 2022).

Susha, I., Grönlund, Å. and Van Tulder, R. (2019) 'Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions', *Government Information Quarterly*, 36(1), pp. 112–128.

Susha, I., Janssen, M. and Verhulst, S. (2017) 'Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development', in T. Bui (ed.) *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1–10.

Taddeo, M. (2017) 'Data philanthropy and individual rights', *Minds and Machines*, 27(1), pp. 1–5.

Taylor, P.L. and Mandl, K.D. (2015) 'Leaping the data chasm: structuring donation of clinical data for healthcare innovation and modeling', *Harvard health policy review: a student publication of the Harvard Interfaculty Initiative in Health Policy*, 14(2), p. 18.

Tenopir, C. *et al.* (2011) 'Data Sharing by Scientists: Practices and Perceptions', *PLoS ONE*. Edited by C. Neylon, 6(6), p. e21101. Available at: https://doi.org/10.1371/journal.pone.0021101.

Wenger, E. (1998) *Communities of Practice: Learning, Meaning, and Identity (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.

Wixom, B.H., Sebastian, I.M. and Gregory, R.W. (2020) 'Data Sharing 2.0: New Data Sharing, New Value Creation', *MIT CISR* [Preprint].

Yin, R. (2003) *Case Study Research: Design and Methods, Third Edition, Applied Social Research Methods Series, Vol 5*. London, UK: Sage Publications, Inc.