NETWORK LAW REVIEW

ARTICLE

Towards Efficient Data Sharing in Platform Markets

Bertin Martens, Geoffrey Parker, Georgios Petropoulos, and Marshall Van Alstyne

Dear readers, the Network Law Review is delighted to present you with this month's guest article by Bertin Martens, Senior Fellow at Bruegel and Non-resident Research Fellow at the Tilburg Law & Economics Centre (TILEC) at Tilburg University, Geoffrey Parker, Charles E. Hutchinson '68A Professor of Engineering Innovation at Dartmouth College, Georgios Petropoulos, Assistant Professor of Data Sciences and Operations at USC Marshall School of Business, and Marshall Van Alstyne, Allen and Kelli Questrom Professor in Information Systems at Boston University & Harvard Law School Berkman Center.

Abstract: Platforms collect valuable data that they do not share. Welfare then suffers as information asymmetry produces market failures. Current proposals to address this, however, fail because they grant access to an entity's own data but not the context within which it functions. We propose a novel solution, the in-situ data right, that gives users privacy control yet allows third parties to create value with user permission.

***

Data Externalities and Information Asymmetry in Platform Markets

In traditional offline two-sided markets, such as town markets, buyers and sellers gather in a physical place and benefit from number-driven network effects: more buyers attract more sellers, and vice versa. Users collect their own market information to make their transaction decisions.[1] This decentralized information system is economically inefficient for two reasons. First, it is privately inefficient

[1] Granovetter, M. (1973). The Strength of Weak Ties: A Network Theory Revisited. Sociological Theory, Vol 1, p 201.

because information collection costs often restrict the amount of information from which users can benefit. With incomplete information, they are more likely to make inefficient decisions. Second, it is socially wasteful because, in the absence of a data-sharing mechanism, each user must collect the same information as other users. In the absence of digital data collection technology, search friction alone is a standard information market failure, yet there is little that market organizers can do to make markets more transparent and facilitate matching among users.

The arrival of digital technology can overcome these problems. Digital platforms collect and centralize a rich set of market information, including user preferences, interactions, and transaction data.[2] This market data pool is a potential source of at least three types of welfare externalities. First, economies of scope in the reuse of non-rival information reduce the social cost of information collection.[3],[4] Second, economies of scale in the aggregation of user data,[5],[6],[7] enrich the insights that can be derived from a centralized market data pool, as compared to fragmented user data. Third, networked user interaction data offers more insight than disconnected and unrelated individual data. These data externalities increase the social value of pooled market data and improve service quality on digital platforms.[8],[9] Scholars and commentators sometimes label the combination of all these data externalities "data-driven network effects."[10],[11] Data traces left behind by user interactions help platforms to improve service quality and attract more users to the platform. This is different from traditional user-driven network effects, where the number of users alone improves service value, irrespective of their data exhaust. As such, platform

[2] Martens, B. (2021). Data access, consumer interests and social welfare–an economic perspective on data. In Data Access, Consumer Interests and Public Welfare, pp 69–102. Federal Ministry for Justice and Max Planck Institute for Innovation and Competition, editors. Nomos publishers, Baden-Baden, Germany, 2021.

[3] Panzar, J C, R D. Willig (1977) Economies of Scale in Multi-Output Production, The Quarterly Journal of Economics, Volume 91, Issue 3, August 1977, Pages 481–493.

[4] Teece, D (1980) Economies of scope and the scope of the enterprise, Journal of Economic Behavior & Organization, Volume 1, Issue 3, 1980, Pages 223-247.

[5] Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. In AEA Papers and Proceedings, volume 109, pages 33–37.

[6] Calzolari, Giacomo and Cheysson, Anatole and Rovatti, Riccardo (2023) Machine Data: Market and Analytics, January 23, 2023. Available at SSRN.

[7] Carballa-Smichowski Bruno, Néstor Duch-Brown, Seyit Höcük, Pradeep Kumar, Bertin Martens, Joris Mulder and Patricia Prüfer (2022) Economies of scope in data aggregation: evidence from health data, working paper Joint Research Centre of the European Commission, November 2022.

[8] Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. American Economic Journal: Micro-economics, vol 14(4), November 2022, (pp. 218-56).

[9] Choi, Jay Pil, Doh-Shin Jeon and Byung-Cheol Kim (2019) Privacy and personal data collection with information externalities, Journal of Public Economics, Vol 173, pages 113-124.

[10] Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. Academy of management review, 46(3), 534–551.

[11] Prüfer J. and C. Schottmüller (2022) 'Competing with big data', The Journal of Industrial Economics, Vol LXIX(4).

data collection can, in principle, make markets more transparent and enable more efficient user decision making.

In practice, however, platforms do not share their full market information with users. Profit-maximizing platform operators want to retain exclusive data control to maximize revenue from data-driven matching services that they offer to users.[12] Full disclosure of all platform information would undermine data monetization strategies and weaken incentives to collect data. Instead, profit-maximizing platforms share only narrow market information signals with their users, through paid and organic recommendations, especially when they are vertically integrated in the markets they operate. Although platform market entry pricing still leaves a non-monetized user surplus,[13] biased platform signals and remaining information asymmetries between platforms and users result in an information market failure. Users benefit only partially from the data externalities that they contribute to and leave behind on a platform. There are significant margins for more efficient use of market information and increasing user surplus.[14],[15] For example, inside many e-commerce platforms, sellers only observe their own sales but receive no user interaction data with competing products. This results in poorly informed business decisions about product development, market positioning, and pricing, especially when the platform operator is vertically integrated and uses its exclusive information to distort competition. Such asymmetric distribution of information within a platform network can lead to market failures that reduce consumer welfare.

Information asymmetry is particularly costly when the platform is vertically integrated with a business use case.[16] Under vertical integration, the platform has incentives to foreclose the upstream market against third-party providers of goods and services that are in direct competition with the vertically integrated business. Information rents distort competition. So, there is a scope for introducing a data sharing mechanism that requires the platform operator to share market data with its business partners, reducing information asymmetry. Symmetrical information structures within a platform network can resolve market failures and maximize consumer welfare.

---

[12] Bergemann D and A Bonatti (2019) Markets for Information: An Introduction, Annual Review of Economics, Vol. 11:85-107, Volume publication date August 2019.

[13] Brynjolfsson, E., A. Collis and F. Eggers (2019) 'Using massive online choice experiments to measure changes in well-being', PNAS116(15): 7250-7255.

[14] Ursu, R M (2019) The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions, Marketing Science Vol. 37, No. 4, June 2019.

[15] De los Santos, B., & Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. Marketing Science, Vol 36(4), pp 542–564.

[16] Martens, Bertin and Parker, Geoffrey and Petropoulos, Georgios and Van Alstyne, Marshall W., Towards Efficient Information Sharing in Network Markets (January 3, 2024). Proceedings of the 57th Hawaii International Conference on System Sciences 2024, Available at SSRN: http://dx.doi.org/10.2139/ssrn.3954932.

Rents from information bottlenecks occur not only within a platform but also between platforms.[17] Market information collected by one platform can be useful to improve services on another and may increase competition when network effects lock users into a single monopolistic service provider. To some extent, this already happens when platforms become part of intra-platform ecosystems of complementary service providers,[18],[19] or even in competing inter-platform ecosystems.[20] However, strategic decisions on data-driven co-opetition are made in the private interests of competing firms that are not necessarily aligned with user welfare. For example, a platform's private incentives to share data with competitors are, in most cases, lower than the social ones. So, again, the imposition of a data sharing mechanism between platforms could allow competing platforms, with an information disadvantage, to improve their value proposition and attract more users. More symmetric data access would enable more effective competition at the intermediary level, increasing consumer choice and welfare.

The Challenge of Data Governance and Existing Regulations

Going from theory to practice, technical protection measures give platforms de facto exclusive control over the data they collect. Exclusive control stands at odds with the fact that platform data is co-generated between at least three, and possibly more, parties: two interacting parties on the platform, and the platform operator that enables this interaction. All these parties may have claims to access and make alternative uses of the data. Exclusive control by a single party is unlikely to maximize the social value of platform data. Making efficient use of all potential data externalities requires governance rules, possibly a combination of private governance and regulatory intervention[21] that assigns a variety of access rights to different parties.

Regulators are increasingly aware of data market failures within and between platforms. They have already granted users access and portability rights to their personal data in the EU General Data Protection Regulation (GDPR)[22] and in the

[17] Economides, N., & Lianos, I. (2021). Restrictions on privacy and exploitation in the digital economy: a market failure perspective. Journal of Competition Law & Economics, 17(4), 765-847.

[18] Jacobides, M. G., Sundararajan, A., and Van Alstyne, M. (2019). Platforms and ecosystems: Enabling the digital economy. In World Economic Forum briefing paper.

[19] Jacobides, M. G., Cennamo, C., and Gawer, A. (2021). Distinguishing between platforms and ecosystems: complementarities, value creation and coordination mechanisms. Working paper.

[20] Hannah, D. and K. Eisenhardt (2018) How firms navigate cooperation and competition in nascent ecosystems, Strategic Management Journal, Vol 39(12), December 2018, pages 3163-3192.

[21] Ottolia, A and C Sappa (2022) A topography of data commons: from regulation to private dynamism, GRUR International, vol 71(4), April 2022, pp 335-345.

[22] European Union (2016) Regulation (EU) 2016/679 of the European Council and the Parliament of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation, GDPR).

California Consumer Privacy Act (CCPA).[23] Portability rights were rendered more operational in the EU Digital Markets Act (DMA)[24] and expanded to business users' commercial data, at least in very large "gatekeeper" platforms. However, platforms collect user interaction data that goes beyond their "own" data. The DMA introduced some case-specific provisions for sharing network interaction data, for example, search engine data. Other EU data regulations are gradually introducing access to user interaction data, for example, in health data[25] and other industrial data pooling initiatives.[26] The DMA is primarily designed as a competition policy instrument to overcome the negative welfare consequences of network effects that can "tip" markets towards a single dominant platform.[27],[28],[29],[30] It acknowledges that data may further entrench monopolistic market positions.[3132] Data sharing obligations are perceived as a means to overcome anti-competitive harms of network effects.

The "In-Situ" Data Right: A Path Forward

Our objective here is to go a step further and propose an effective network data sharing policy that complements this competition policy framework to maximize the social benefits from regulatory intervention. The solution should, in principle, (i) bridge the gap between the realized private value and the potential social value of data, (ii) maintain the economic vitality and viability of platforms, and (iii) respect user privacy rights. This requires enhancing and expanding existing data access regulations. While sharing "own" data may generate economies of scope in the re-use of non-rival data, we can explore several cases where sharing of networked interaction data can level the information level playing field and further increase welfare through economies of scale and scope in data aggregation.

[23] State of California (2018) California Consumer Privacy Act.

[24] European Union (2022) Regulation (EU) 2022/1925 of the European Parliament and the Council of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act).

[25] European Commission (2022) Proposal for a Regulation on the European Health Data Space, COM(2022) 197 final, May 2022.

[26] European Commission (2020) A European Strategy for Data, Communication from the European Commission to the Council and the Parliament, COM(2020)66 Final, 19 February 2020.

[27] Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T. M., & Van Alstyne, M. W. (2021). The EU Digital Markets Act: A report from a panel of economic experts. Publications Office of the European Union, Luxembourg.

[28] Crémer, J, Y-A de Montjoye and H Schweitzer (2019) Competition policy for the digital age, Directorate General for Competition, European Commission, Brussels.

[29] Furman, J., Coyle, D., Fletcher, A., McAuley, D., & Marsden, P. (2019). Unlocking digital Competition, report of the digital competition expert panel (Tech. Rep.). UK Government.

[30] Scott Morton, F. (2019). Final report of the subcommittee on market structure and antitrust(Tech.Rep.). Stigler Committee on Digital Platforms.

[31] Prüfer J. and C. Schottmüller (2022) 'Competing with big data', The Journal of Industrial Economics, Vol LXIX(4).

[32] Hagiu, A., and Wright, J. (2023). Data-enabled learning, network effects and competitive advantage. Rand Journal of Economics, Fall 2023.

At the same time, pre-existing exclusive rights to data can complicate data sharing. For example, the EU GDPR grants natural persons exclusive rights to their personal data.[33] Under the EU Trade Secrets Directive,[34] sellers can claim exclusive rights to some of their commercial data. Platforms may claim exclusive rights to inferential or processed user interaction data because they were facilitated by the platform infrastructure and algorithms. Pre-existing rights may be an obstacle to accessing the networked data of individual users, unless rights holders consent to their re-use, which may be difficult and costly to obtain. Access to aggregated data can sometimes be enough to overcome this obstacle.

When access to individual network interaction data is required to overcome information asymmetry, we propose introducing an "in-situ" data right, whereby business users may import algorithms of their choice into the infrastructure where their data are resident.[35] Users, or third parties acting on their behalf, can then derive insights from their data and its context in a secure setting. Through tools such as Federated Learning technology, such analysis need not reveal the underlying primary data.[36] This narrows the gap between the private and social value of data.

Implementing In-Situ Access: Second-Degree Data and Privacy

We propose a two-step approach to an economically meaningful interpretation of "making available":

1. A new data right for business users to gain access to individual and disaggregated second-degree network interaction data between buyers and sellers. This should make it possible for users to estimate distances to their nearest competitors.
2. Consumer privacy protection can be ensured through in-situ data access: sellers can bring their algorithms to the data inside the platform, rather than taking the data ex-situ out of the platform.

DMA Article 6 §10 already mandates that platforms share user interaction data with business users. These are zero-degree direct interaction data (sales and views) with consumers (see Fig. 1). First-degree interaction data would include information

[33] European Union (2016) Regulation (EU) 2016/679 of the European Council and the Parliament of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation, GDPR).

[34] European Union (2016) Directive (EU)2016/943 of the European Council and the Parliament of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against unlawful acquisition, use and disclosure.

[35] Van Alstyne, M. W., Petropoulos, G., Parker, G., & Martens, B. (2021). 'In situ' data rights. Communications of the ACM, 64 (12), 34-35.

[36] Ramírez, D.H., Díaz, L.P., Rahimian, S., García, J.M.A., Peña, B.I., Al-Khazraji, Y., Alarcón, Á.J.G., Fuente, P.G., Soler Garrido, J. and Kotsev, A.(2023) Technological Enablers for Privacy Preserving Data Sharing and Analysis, Joint Research Centre of the European Commission.

on consumers who interact with a seller's products and that seller's competitors. Buyers use information intermediaries across two sellers. Second-degree interaction data involves two intermediaries: a buyer and another seller. This includes data from consumers who interacted with products from competing sellers but did not buy from one of them. If available, such data gives the first seller insight into the preferences of buyers who browsed and bought from their competitors and the characteristics of competing products. The combination of these networked interaction data enables sellers to assess consumer preferences and willingness-to-pay for product characteristics for a range of close-substitute products, including prices and sales estimates for these products. This provides sufficient information to estimate distances to the nearest competitors within a given product category and (re-)position toward optimality in the platform market, in line with the full information scenario above.

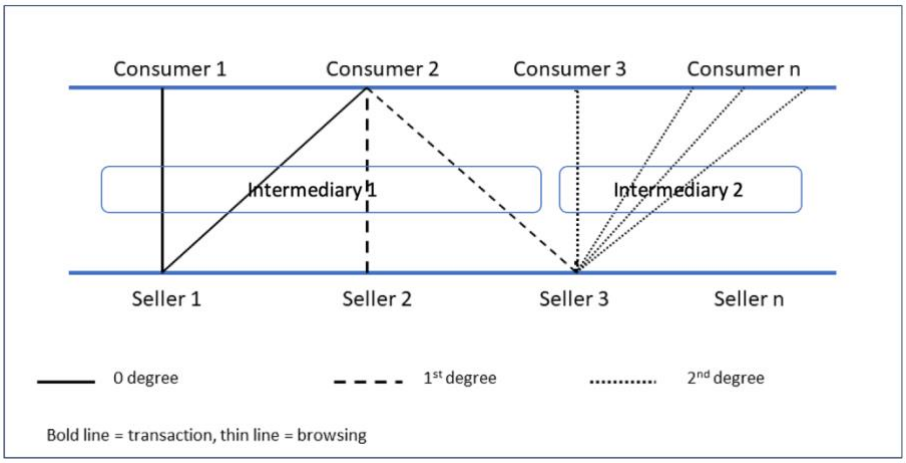


Figure 1: Degrees of interaction, starting from a 0-degree direct link between Seller 1 and Consumer 1

With access to second-degree interaction data, not all platform data is shared with its business users. However, it reduces the platform's information advantage compared to zero-degree data access alone. Preserving part of the information asymmetry is important not to undermine the economic viability of the platform as a central intermediary. Putting all platform information into the public domain would undermine the platform's incentive to invest as well as its commercial viability, putting at risk the positive social network effects that it generates for its users. Exclusive control over at least part of its data pool facilitates monetization of the data through advertising.

Second-degree networked data includes access to buyers' interactions and profile data. In case buyers are individuals, this constitutes personal data subject to privacy protection under the EU GDPR or California's state-sponsored CCPA. It can only be accessed with the consent of the buyer or in an anonymous way, as emphasized by

DMA Article 6 §10. It is costly and unlikely that sellers will obtain consent from a sufficiently large number of first-degree buyers, let alone a wider network of second-degree interactions. This leaves anonymous data access as the only viable option. This can be achieved through user data aggregation. However, that may destroy valuable context information and increase error margins in the estimation of distances to nearest competitors. A more information-efficient option is in-situ data access that preserves detailed user information. Rather than taking data out of the secure space of the data holder or platform operator and porting it ex-situ to the seller's infrastructure, in-situ reverses the operation and brings seller algorithms to the data holder's secure space for in-situ computations. Sharing networked information ex-situ increases the risk of de-anonymization of personal data[37] and violation of data privacy rules. Under in-situ, the data holder can monitor computations and restrict information output to derived data while protecting the detailed input data. Several computational technologies exist today that facilitate secure and privacy-preserving in-situ computation. For example, Ramírez, et. al.[38] find that federated learning performs better than secure multi-party computation, differential privacy, and homomorphic encryption, provided there is sufficient computational capacity and data coherence for the computational model to converge. Smaller business users on a platform may require help to perform the necessary data analytics, from a third-party provider or from the platform itself. Platforms will no longer be in a position to charge a monopolistic price for such services. Pricing of in-situ use of the platform computing infrastructure should be regulated at marginal cost, as part of the in-situ access right.

Conclusion

Sharing network interaction data matters for the efficient use of platform data. While current regulations represent a step in the right direction, they fall short of requirements for efficient data use. In particular, they fail to provide sufficient access to platforms' networked interaction data. Mandatory data sharing provisions will need to be adjusted and expanded to ensure a more efficient balance between competition concerns and data-driven network effects. In situ data rights move the bottleneck from the platform to the users, who can capture a greater share of the value their data creates. These rights retain context so more value can be created, and they foster fair competition by improving symmetry of access among competing platforms and startups.

[37] De Montjoye, Y.-A., Radaelli, L., Singh, V. K., & Pentland, A. (2015). Unique in the shopping mall: On the re-identifiability of credit card metadata. Science, 347(6221), 536–539.

[38] Ramírez, D.H., Díaz, L.P., Rahimian, S., García, J.M.A., Peña, B.I., Al-Khazraji, Y., Alarcón, Á.J.G., Fuente, P.G., Soler Garrido, J. and Kotsev, A.(2023) Technological Enablers for Privacy Preserving Data Sharing and Analysis, Joint Research Centre of the European Commission.

Bertin Martens, Geoffrey Parker,
Georgios Petropoulos, and Marshall Van Alstyne

Author Disclosures:

Geoffrey Parker testified in Federal Tax Court for the Internal Revenue Services in its lawsuit against Facebook in Docket No. 21959-16. The testimony took place in March 2022.

Marshall Van Alstyne provided expert advice to Meta on antitrust concerns of Facebook Marketplace relative to Craigslist in August 2022. He declined compensation to avoid a conflict of interest.