DATA BROKERS:
INTERMEDIARIES FOR MORE EFFICIENT DATA MARKETS?

Respectively, Chair of Machine Learning and Uncertainty Quantification, University of Regensburg, andreas.schauer@ur.de and Chair of Machine Learning and Uncertainty Quantification, University of Regensburg, daniel.schnurr@ur.de.
INTRODUCTION

By now, data has become an important economic resource for firms and organizations to create manifold business value across diverse use cases and application domains. In consequence, data has turned into an economic good that can be shared and traded between organizations and individuals. According to the Data Market Study 2021-2023, the European data market is currently growing at an annual rate of 12.6 percent and has reached €72.9 billion in the European Union in 2022. About one quarter of this market value can be attributed to the monetization of data. However, data exhibits several peculiar characteristics as an economic good, which may impede the emergence and well-functioning of data markets and therefore present overarching policy challenges with regard to data fragmentation and data concentration.

Data brokers play a pivotal role in addressing both these challenges by acting as intermediaries that facilitate data sharing between organizations and also individuals. By collecting, aggregating, enriching, and exchanging both personal and non-personal data, data brokers are envisioned to improve the access to data for a broader range of organizations. In consequence, even firms that do not have direct access to data sources could benefit from data-driven business models, data-driven quality improvements, and data-driven innovations. In this vein, data brokers could alleviate concerns about data concentration. This applies in particular to user data that is generated as a by-product during consumers' usage of a digital service and therefore entails data-driven network effects. At the same time, data brokers can mitigate data fragmentation by establishing specialized institutions, trading mechanisms, and economic incentives for data sharing. This can generate social benefits and provide new opportunities for innovation through the sharing, aggregation, and combination of otherwise isolated and fragmented complementary data sets. Finally, personal data brokers are envisioned to offer immediate benefits to individuals and data subjects so that they can reap some of the business value associated with the sharing and processing of their data.

However, there are also widespread concerns about the business practices of current data brokers and the general transparency of the market for data. This applies especially to data brokers that collect, process, and share personal data, often without the knowledge and explicit consent of the data subjects. Individuals often face difficulties in effectively exercising their rights to privacy and informational self-determination due to the inherent information asymmetries that face them vis-à-vis data brokers. This can yield significant privacy risks and may undermine individuals' trust in the data economy in general. Even if individual data sets are anonymized or pseudonymized, combining granular data can inadvertently reveal sensitive information or personal identities. In consequence, this presents novel challenges in ensuring privacy for the sharing of granular behavioral data. In the context of non-personal data, similar concerns about data security and confidentiality can arise if data sets reveal commercially sensitive data, such as trade secrets.

In this article, we review and summarize key strands of the academic literature on data brokers and highlight the challenges for data markets and data intermediaries that arise from the special characteristics of data as an economic good. We then highlight recent findings on the economic impact of data brokers in digital markets and scrutinize the concept of personal data. In this article, we review and summarize key strands of the academic literature on data brokers and highlight challenges for data markets and data brokers that arise from the special characteristics of data as an economic good. We then highlight recent findings on the economic impact of data brokers in digital markets and scrutinize the concept of personal data.

7 See, e.g., FTC, supra note 5.
8 Yves-Alexandre De Montjoye, Laura Radaelli, Vivek Kumar Singh & Alex “Sandy” Pentland, Unique in the shopping mall: On the reiden-
In general, data brokers can be distinguished based on the type of data they focus on. Business-to-business (“B2B”) data brokers primarily facilitate the exchange of information between firms and organizations, for example by operating industrial data platforms. In contrast, personal data brokers focus on data related to individual consumers, often including personal information. While consumers may have the opportunity to voluntarily share their own personal data with these intermediaries, personal data brokers commonly rely on two main sources: acquiring data from private enterprises and government agencies as well as actively collecting data from private enterprises. While consumers may have the opportunity to voluntarily share their own personal data with these intermediaries, personal data brokers commonly rely on two main sources: acquiring data from private enterprises and government agencies as well as actively collecting data from private enterprises. Therefore, these data brokers frequently collect, combine, analyze, and monetize individuals’ data without the explicit knowledge of the data subjects. Although such data brokerage can nonetheless yield benefits for consumers (for example, by preventing fraudulent activities, enhancing product offerings, or enabling more informative personalized advertisements), it is also associated with inherent risks, particularly concerning consumer privacy.11

A. The Special Characteristics of Data as an Economic Good and Associated Challenges for Data Markets

As data serves as a valuable input for diverse applications and digital services, data has become an economic good in itself that can be priced and traded between organizations. However, data exhibits several special characteristics that can present barriers to the emergence of efficient data markets.12

1. Non-rivalry of data: Data can be used by multiple parties simultaneously for different purposes without it being depleted or diminishing its original quality and functionality. From a welfare perspective, non-rivalry implies that data should be shared and used widely among firms and organizations to maximize social benefits.13 However, non-rivalry may also discourage firms from sharing and selling data, when they fear that sharing the data with other firms could strengthen potential competitors and increase the risk of creative destruction. Moreover, non-rivalry may limit the exclusive use of data and thus limit firms’ ability to profitably sell their data (see Section II.B.).

2. Data quality: Data is a heterogeneous product and subject to quality differentiation,14 which can be measured along multiple dimensions, such as accuracy, completeness, timeliness, or consistency.15 Moreover, the information quality, which denotes the “fitness for use” of a particular data set and determines the ultimate business value of the data in a specific use case, is highly context-dependent and subjective. Therefore, the same set of data can be of very different value to different firms and organizations, which complicates the commodification of data and the emergence of data markets.

3. Data as an experience good and the Arrow information paradox: The pricing and trading of data is further complicated by data being an experience good, meaning its true quality and value can only be determined by an organization after it has acquired or used the data. This leads to the famous Arrow information paradox16, according to which a buyer must gain detailed knowledge about the information in a data set to assess its value. However, after having established this knowledge, there is no more incentive for the buyer to acquire the data. In consequence, this can lead to sustained information asymmetries in data markets, which can prevent further barriers to trade and efficient market outcomes.

4. Data-driven network effects: When data is created as a by-product of usage, this can give rise to data-driven network effects. For example, the more data a firm can use to improve the quality of its service, the more users will be attracted by the service, which in turn generates more data.18 In consequence, data-driven network effects can propel positive feedback loops and promote market concentration, as data-rich firms take over an increasing share of a market, which can further hamper incentives to share data.

5. Economies of scale and scope: Economies of scale and scope in data collection and data use may further promote data concentration.19 In addition, there are often scale and scope advantages from combining the utilization of the necessary technical infrastructure, algorithms, and skilled employees. These market characteristics can create additional barriers to the free flow of data.

To overcome these challenges, B2B data brokers may develop and establish economic institutions to facilitate the sharing and trading of non-personal data. Recent analytical and empirical studies explore such market mechanisms and investigate the effects of control and transparency on firms’ incentives to share data and the well-functioning of data markets. Rasouli et al. propose market mechanisms and optimal pricing schemes for sharing data against money as well as sharing data against data.20 These mechanisms leverage firms’ ability to artificially alter the quality of the shared data to achieve socially optimal outcomes. In an experimental study on B2B data-sharing platforms, Krämer et al. demonstrate that giving firms on such platforms control over which other firms can access their data promotes data sharing.21 The same holds for increasing transparency about other firms’ decisions to share data and the resulting data transactions on the platform. This is because increased control and transparency allow firms to punish and deter other firms from free-riding, hence creating an incentive for participating in data sharing.

B. Economic Impact of Data Brokers as Intermediaries

Several recent studies investigate the role of data brokers as information intermediaries in digital markets. To this end, various game-theoretic analyses examine the optimal strategies and the economic impact of monopolistic data brokers that sell consumer information to retailers in downstream markets, where the data is used for better demand
to merge their individual data sets and sell a joint data set. Therefore, the peculiar characteristics of data as an economic good have important implications for data brokers’ incentives to differentiate their data sources and the competition intensity in data markets.

**Insight 2: Data brokers can promote competition in downstream markets that make use of the data. However, it is often optimal for data brokers to offer data exclusively or to differentiate its quality when selling to downstream firms. There is also the risk that data brokerage may reinforce existing data concentration. The non-rivalry of data softens competition between data brokers and may lead to lower rewards for data creators.**

For competing data brokers, Ichihashi highlights that the non-rivalry of data reduces intermediaries’ incentives to offer consumers a high reward for their data, as consumers may sell the same data to other intermediaries, thus lowering the commercial value of the data. Thus, non-rivalry softens competition and lowers the reward for data creators. On a related note, Ichihashi as well as Gu et al. study rivalry softens competition and lowers the reward for data creators. Gu et al. find that a data broker can increase its profit by selling more than one downstream firm, but only if it supplies data of different qualities to different retailers. Similarly, Bounie et al. show that a data broker can profitably split its data set and sell mutually exclusive partitions to different downstream firms. In both cases, the distinct input data sets allow retailers to differentiate themselves, which softens competition with personalized prices in the downstream market and thus increases retailers’ willingness to pay for the data. For competing data brokers, Ichihashi highlights that the non-rivalry of data reduces intermediaries’ incentives to offer consumers a high reward for their data, as consumers may sell the same data to other intermediaries, thus lowering the commercial value of the data. Thus, non-rivalry softens competition and lowers the reward for data creators. On a related note, Ichihashi as well as Gu et al. demonstrate that competing data brokers have an incentive to collect and sell exclusive data sets to evade the negative impact of intense competition or may even prefer to merge their individual data sets and sell a joint data set. Therefore, the peculiar characteristics of data as an economic good have important implications for data brokers’ incentives to differentiate their data sources and the competition intensity in data markets.

**Monets et al. demonstrate that it is optimal for monopolistic data brokers to rely on exclusive contracts when selling a single data set of consumer information.**

**C. Privacy Risks and Consumer Empowerment**

A further stream of empirical literature investigates the practices of personal data brokers in today’s data economy and highlights the potential threats to data subjects that can emerge from these practices. In particular, data brokers lack of transparency in collecting and managing individual’s information often undermines consumers’ data control, posing significant privacy risks. As highlighted by the FTC, the indefinite retention periods practiced by data brokers present additional privacy and security risks. Neumann et al. find that user profiles offered for sale by data brokers are frequently of low data quality, limiting their economic value and raising further privacy concerns. In particular, the study evaluates data brokers’ ability to infer demographic information and interests, revealing frequent inaccuracies, especially in gender prediction. In a survey on 75 data brokers, Ameur et al. show that personal data brokers offering personal information for free on their websites can reveal sensitive personal information and pose significant risks to data subjects, as data from different sources can easily be linked across different data brokers by malicious actors using data matching techniques.

Building on these findings, recent studies analyze approaches to make brokerage of personal data more privacy-preserving and to involve data subjects more directly in the sharing of their data. From a technical perspective, privacy-preserving technologies, such as the encryption and signing of data to ensure identity preservation, data integrity, and data confidentiality, have been proposed to mitigate the risks of data sharing and to increase data subjects’ trust in data markets. Moreover, blockchain-based infrastructures have been suggested to facilitate user-controlled data sharing, although their effect on privacy is controversial.

**From a policy perspective, the literature documents recent efforts by policymakers to give consumers more power and control over the sharing and monetization of their personal data, as well as the European Commission’s emphasis that Personal Information Management Services (“PIMS”) could serve as a key building block for a user-centric data economy. In particular, PIMS allow data subjects to store, manage, and share data under their own control and could thus serve as a technical infrastructure to support business models that allow consumers to sell their data in return for a monetary reward. To this end, it is important that data subjects can retrieve and aggregate their personal data from digital services that they regularly use. In the European Union, this is supported and facilitated by the right to data portability under the General Data Protection Regulation ("GDPR"). Technically, data transactions through personal data brokers could further be supported by smart contracts and blockchain-based infrastructures.**

**Monets et al. demonstrate that it is optimal for monopolistic data brokers to rely on exclusive contracts when selling a single data set of consumer information.**

**Building on these findings, recent studies analyze approaches to make brokerage of personal data more privacy-preserving and to involve data subjects more directly in the sharing of their data.**

**References:**


24. Xin Zhang, Wei Thoo Yue, Yugang Yu & Xiong Zhang, note 26.

25. C. Privacy Risks and Consumer Empowerment

A further stream of empirical literature investigates the practices of personal data brokers in today’s data economy and highlights the potential threats to data subjects that can emerge from these practices. In particular, data brokers lack of transparency in collecting and managing individual’s information often undermines consumers’ data control, posing significant privacy risks. As highlighted by the FTC, the indefinite retention periods practiced by data brokers present additional privacy and security risks. Neumann et al. find that user profiles offered for sale by data brokers are frequently of low data quality, limiting their economic value and raising further privacy concerns. In particular, the study evaluates data brokers’ ability to infer demographic information and interests, revealing frequent inaccuracies, especially in gender prediction. In a survey on 75 data brokers, Ameur et al. show that personal data brokers offering personal information for free on their websites can reveal sensitive personal information and pose significant risks to data subjects, as data from different sources can easily be linked across different data brokers by malicious actors using data matching techniques.

Building on these findings, recent studies analyze approaches to make brokerage of personal data more privacy-preserving and to involve data subjects more directly in the sharing of their data. From a technical perspective, privacy-preserving technologies, such as the encryption and signing of data to ensure identity preservation, data integrity, and data confidentiality, have been proposed to mitigate the risks of data sharing and to increase data subjects’ trust in data markets. Moreover, blockchain-based infrastructures have been suggested to facilitate user-controlled data sharing, although their effect on privacy is controversial.

**From a policy perspective, the literature documents recent efforts by policymakers to give consumers more power and control over the sharing and monetization of their personal data, as well as the European Commission’s emphasis that Personal Information Management Services (“PIMS”) could serve as a key building block for a user-centric data economy. In particular, PIMS allow data subjects to store, manage, and share data under their own control and could thus serve as a technical infrastructure to support business models that allow consumers to sell their data in return for a monetary reward. To this end, it is important that data subjects can retrieve and aggregate their personal data from digital services that they regularly use. In the European Union, this is supported and facilitated by the right to data portability under the General Data Protection Regulation ("GDPR"). Technically, data transactions through personal data brokers could further be supported by smart contracts and blockchain-based infrastructures.**

**References:**


33. FTC, supra note 5.


42. Mattias Tavzanio, Martin Minnori, Gustavo Azperrua, Carlos Sarauate & Nicolas Della Penna, Watson: A decentralized marketplace empowering individuals to safely monetize their personal data (White Paper, 2018); Xin Zhang, Wei Thoo Yue, Yugang Yu & Xiong Zhang, supra note 26.

43. Sachit Mahajan, Data Marketplaces: A Solution for Personal Data Control and Ownership?, Sustainability 14(24), 16884 (2022).

In practice, there have been numerous examples of startup firms like Datacy45 and Datum46 promising to establish such personal data brokers to the benefit of consumers. However, so far, the success of these businesses has been very limited, with several personal data brokers having stopped their operations without paying significant rewards to consumers.47 Haberer et al. show that this may be explained by the strategic interactions and the economic effects that arise in the context of personal data brokers.48 In particular, providers of digital services, where the data is generally of high quality and usage, may invest less into the quality of their services when a personal data broker competes for the same data-driven revenues. This is especially the case when monetary rewards do not remunerate consumers for their existing data, but also offer incentives for consumers to create additional data through more usage. Only if the provider of the digital service can appropriate some of the consumers’ rewards from the personal data broker through a higher price for its service, the provider will be willing to raise the quality of its service again. In consequence, consumers can only benefit from personal data brokers if these brokers are very efficient in generating revenues from data and can therefore increase the industry’s overall data revenues. Evidently, current personal data brokers seem not to meet this efficiency threshold and thus can only sustain a marginal existence where they pay only negligible rewards to consumers. This further questions whether personal data brokers can fulfill the expectations of European policymakers to serve as a core building block of a user-centric data economy.

In practice, there have been numerous examples of startup firms like Datacy and Datum promising to establish such personal data brokers to the benefit of consumers.

**Insight 3: Lack of transparency, indefinite data retention, risks of malicious access and data leakage, as well as the disclosure of sensitive information from data combination pose significant privacy risks for data subjects in the context of personal data brokers. In combination with PIMS, personal data brokers can empower consumers and may allow data subjects to directly sell their personal data. However, consumers can also be worse off with personal data brokers if these brokers are not sufficiently efficient in generating data revenues.**

The GDPR has established a general regulatory framework for the brokerage and sharing of personal data in the European Union. To this end, the GDPR has strengthened the rights and control of data subjects, which may limit the business opportunities and practices of data brokers. However, with the right to data portability, the GDPR has also established new rules that can foster the availability and free flow of data. Based on a game-theoretic analysis, Ke and Sudhir conclude that the privacy rights stipulated by the GDPR are likely to reduce the total volume of consumer data available in the market, whereas the overall effect on consumer welfare depends crucially on the competition intensity in a market.49 In practice, empirical studies show that the right to data portability has so far had a very limited impact in unlocking personal data for effective data sharing, which is commonly attributed to technical challenges and the limited scope of data access.50

In this context, the recently adopted Data Act provides an extended right to data portability to both consumers and business users and extends the scope of data access to connected products and related digital services.51 Under the Data Act, users may not only access data themselves but can also request that data holders transfer the data directly to an authorized third party. This is done with the goal of unlocking new data sources and promoting the free flow of data to mitigate data fragmentation. However, only the third parties’ data access, such as the need for contractual agreements, compensation based on FRAND (i.e. fair, reasonable and non-discriminatory) terms, or restrictions on the use of the accessed data. These restrictions are intended to safeguard the legitimate interests of data holders but run the risk of undermining the ultimate effectiveness of the Data Act in achieving its goal to increase data availability.52 In particular, these restrictions make it unlikely that data brokers can make effective use of the data that could in principle become available through the new data rights in the Data Act.53 This is exacerbated by further limits on commercial practices, such as the prohibition of exclusive contracting between a user and a third party. Thus, there is the risk that the Data Act will fail short in achieving its intended goals as its rules do not sufficiently account for the role that data brokers and data markets could play in promoting data sharing and the widespread use of data.

As a further key pillar of the European Strategy for Data, the Data Governance Act aims to increase trust in data sharing and overall data availability.54 The rules, which are applicable since September 2023,55 directly address data brokers as key actors of the data economy and impose several requirements for data intermediation services. Most notably, the Data Governance Act requires data intermediation services to notify a competent public authority about its operation, stipulates structural unbundling of intermediation services and prohibits the re-use of data for any purpose other than data intermediation, requires prices for data access to be based on FRAND terms, and imposes additional transparency requirements.56 These requirements for data intermediaries can be expected to have some positive effects on the general transparency of data markets and could help regulators to obtain better information about the identities and practices of data brokers in the market. Furthermore, the legislation establishes a legal basis for regulatory intervention to remedy non-compliance and potential market failures, which could lead to better protection of data subjects. However, the Data Governance Act does not address any of the barriers to data brokerage that stem from the special characteristics of data as an economic good, as identified in this article. Instead, by imposing additional requirements and obligations on data intermediation, the regulation is further diminishing the economic incentives to engage in data brokerage and increases the costs for establishing such businesses. For example, imposing FRAND as a general principle for data access pricing is in stark contradiction to the optimal strategies for data brokers as identified by the economic literature. Overall, the Data Governance Act is therefore more likely to discourage rather than to promote an active data broker industry that could facilitate the free flow of data and increase data availability in the European Union.57

53 id.
54 European Commission, REGULATION (EU) 2022/888 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2022 on establishing a framework for fair access to and use of data (Data Governance Act), L 152/1 (2022).
55 The rules on data intermediation services will become applicable only by September 2025.
CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI’s global community of antitrust experts.