THE GENERATION OF KNOWLEDGE · KNOWLEDGE ECONOMY

Data Marketplaces and the Commercialisation of Knowledge

How knowledge and data are turned into products and traded, what data marketplaces are and why many struggle, the problems of pricing, trust and privacy they must solve, and the rules and models that are shaping how data are shared and sold.

ModQu research article · 2,726 words · 25 references (APA 7th edition)

Knowledge has always been sold in some form: in books, consulting advice, licensed patents and paid research. In recent decades, a new kind of trade has grown up around data. Platforms have emerged to bring together those who hold data and those who want to use them, and governments have begun to write rules for how data may be shared and sold. This article examines how knowledge is commercialised, what data marketplaces are and how they work, why trading data is harder than it looks and what this means for organisations that produce research and data.

Commercialising knowledge

Knowledge can be commercialised in several ways. It can be embodied in products and services, so that customers pay for the product rather than the knowledge itself. It can be protected by patents and copyright and licensed to others. It can be sold as expertise, through consulting and research services. Or it can be packaged as information goods, such as reports, databases and subscriptions.

Arora et al. (2001) studied **markets for technology**, in which firms license patents and know-how to one another. They found that such markets had grown substantially and allowed knowledge to reach firms better placed to use it, but that they worked best where knowledge could be clearly defined and protected. Gans and Stern (2010) asked whether there is a true market for ideas and identified three obstacles. Ideas are hard to value without being disclosed, and disclosure may give them away. Each idea is different, so buyers and sellers struggle to find one another. And ideas can often be copied or resold once acquired. These obstacles, they argued, explain why markets for ideas are thinner and less efficient than markets for most goods.

The first obstacle is a version of the **disclosure paradox** identified by Arrow (1962): a buyer cannot judge the value of information without knowing it, but once the buyer knows it, they no longer need to buy it. Every market in knowledge must find some way around this problem.

An older trade than it seems

Trade in information about people and markets is not new. Credit reporting agencies have compiled and sold information about borrowers since the nineteenth century, allowing lenders to judge the reliability of people they did not know. Financial information services have long sold prices, company accounts and news to investors. Mailing lists of potential customers were bought and sold for direct marketing well before the internet. Market research firms have sold syndicated measurements of retail sales, broadcast audiences and consumer opinion for decades.

What has changed is the scale, speed and variety of data, and the number of organisations that hold data they might sell. Digital activity generates records of purchases, searches, movements and communications in volumes far greater than any earlier source. Sensors in machines, vehicles and buildings add further streams. Computing power allows these data to be combined and analysed cheaply. The question of how such data should be traded, and by whom, has therefore become far more pressing.

What a data marketplace is

A **data marketplace** is an intermediary that connects providers of data with buyers. Spiekermann (2019) described the main types: marketplaces run by independent intermediaries, marketplaces run by a company that sells its own data, exchanges among members of an industry, and personal data platforms through which individuals can control and sometimes sell data about themselves. Figure 1 shows the functions that a marketplace typically performs.

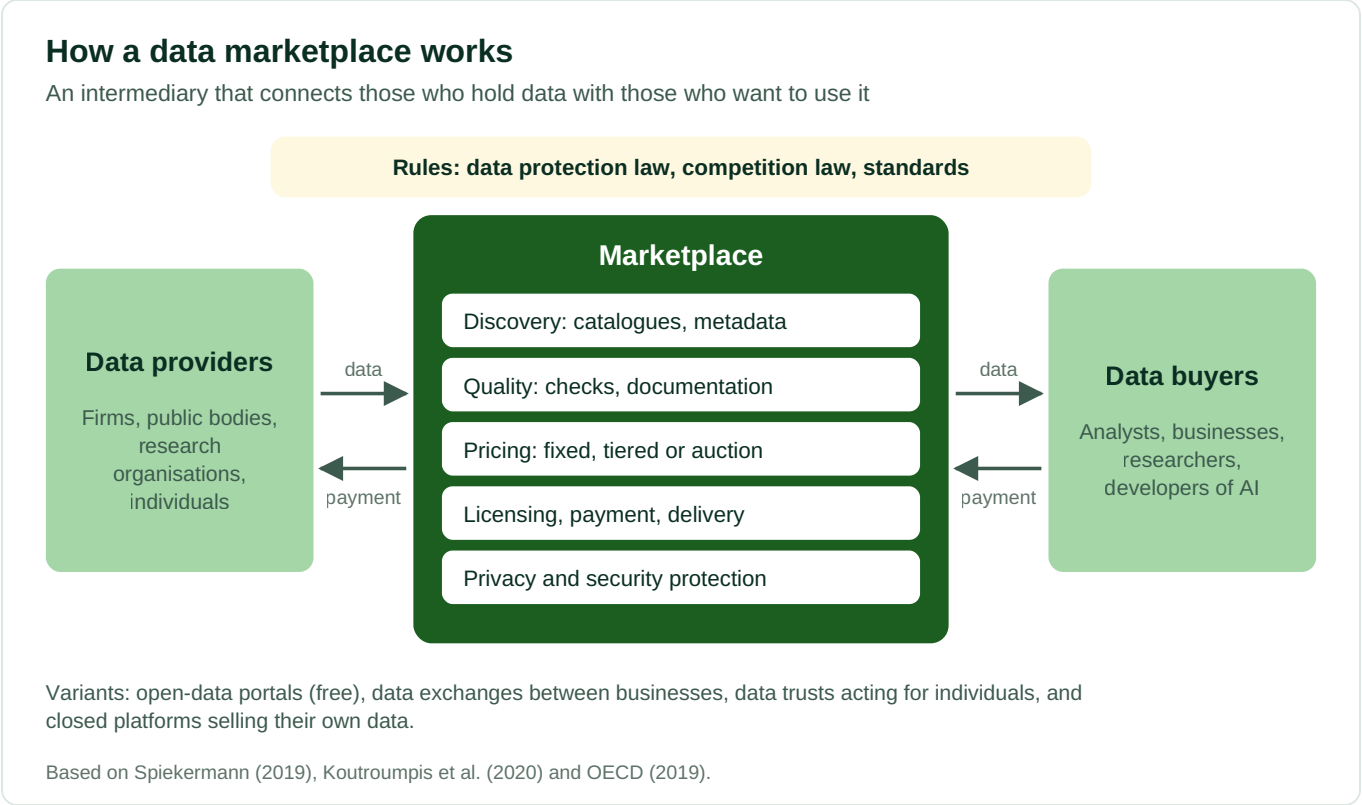


Figure 1. The functions of a data marketplace. Based on Spiekermann (2019), Koutroumpis et al. (2020) and OECD (2019); diagram by ModQu.

Discovery. Buyers need to find data that meet their needs. Marketplaces provide catalogues with descriptions of datasets, their coverage and their structure.

Quality. Because buyers cannot fully inspect data before purchase, marketplaces may check data quality, require documentation and provide samples.

Pricing. Marketplaces set or help set prices, through fixed prices, subscriptions, tiered access or auctions.

Licensing, payment and delivery. Marketplaces handle contracts that define how data may be used, process payments and deliver the data securely.

Privacy and security. Where data concern individuals, marketplaces must ensure that their trade complies with data protection law and protects the people the data describe.

Data marketplaces are a form of **two-sided platform**, which serves two distinct groups of users whose participation depends on each other. Rochet and Tirole (2003) showed that such platforms must attract both sides at once, and that their pricing often charges one side less to attract the other. A data marketplace with few datasets attracts few buyers, and one with few buyers attracts few providers. Parker et al. (2016) described how platforms that solve this problem can grow rapidly, because each additional user makes the platform more valuable to others.

Why trading data is hard

Despite the large volume of data in the world, open data marketplaces have often struggled to grow. Koutroumpis et al. (2020) examined why. They found that most trade in data takes place not on open marketplaces but through bilateral agreements between firms, and they identified several reasons.

The problem of provenance and control

Once data have been sold, the seller loses control over them. A buyer can copy, resell or combine the data in ways the seller did not intend. Koutroumpis et al. (2020) argued that the inability to verify where data came from and to control how they are used is a central barrier to trade. Sellers therefore prefer to deal with buyers they know, under contracts they can enforce, rather than on open markets.

The problem of value

The value of data depends heavily on the buyer's purpose and on what other data the buyer has. Data that are worthless to one buyer may be extremely valuable to another, and the same data may be valuable only in combination with other datasets. This makes it hard to set a price. Bergemann and Bonatti (2019) showed that sellers often do better to sell the results of analysis, tailored to the buyer's decision, than to sell the underlying data. This is one reason why much trade in data takes the form of services rather than transfers of datasets.

Agarwal et al. (2019) proposed a design for a data marketplace in which buyers do not purchase datasets directly. Instead, a buyer submits a prediction task, the marketplace combines data from several sellers to perform it, and the buyer pays according to how much the prediction improves. Sellers are paid according to how much their data contributed. Such designs address the problems of valuation and disclosure by never handing over the raw data, though they remain largely experimental.

The problem of trust

Buyers must trust that data are accurate, complete, lawfully collected and as described. Sellers must trust that buyers will use the data only as agreed. Without trust on both sides, transactions are costly, because each party must check the other. Marketplaces that verify quality and enforce licences can reduce these costs, but only if they themselves are trusted.

The problem of privacy

Much of the most valuable data concerns people: their purchases, movements, health and opinions. Acquisti et al. (2016) reviewed the economics of privacy and found that the effects of sharing personal data are complex. Sharing can create value, for example by improving services and reducing fraud, but it can also harm individuals through discrimination, manipulation and loss of autonomy. Consumers often find it difficult to judge the consequences of sharing data, and their stated concern about privacy does not always match their behaviour.

Removing names and other direct identifiers does not guarantee anonymity. Sweeney (2002) showed that combinations of ordinary attributes, such as postal code, date of birth and sex, were enough to identify a large share of people in the United States, and demonstrated how supposedly anonymous records could be linked to named individuals. Ohm (2010) argued that the failure of anonymisation undermined the legal assumption that de-identified data were safe to share.

New techniques aim to allow data to be used while protecting individuals. **Differential privacy** adds carefully calibrated random noise to statistics or analyses so that the presence or absence of any single person cannot be detected, while preserving the accuracy of results for groups (Dwork & Roth, 2014). Abowd (2018) described the decision of the United States Census Bureau to adopt differential privacy for the 2020 census, reflecting the view that traditional methods of protecting confidentiality could no longer resist modern methods of re-identification.

How data are priced

Data marketplaces use several pricing models. Some charge a fixed price for a dataset or a fixed subscription for continuing access. Others charge by volume, for example per record or per query. Some use tiered pricing, offering basic data cheaply and more detailed, more frequent or more exclusive data at higher prices. A few use auctions. Shapiro and Varian (1999) showed that such strategies follow from the economics of information goods: because the cost of supplying an extra copy is close to zero, prices must be set according to the value to different buyers, and offering several versions allows buyers to sort themselves by how much they value the data.

Exclusivity has a price of its own. Data sold to one buyer only are worth more to that buyer, because competitors cannot use them, but the seller forgoes revenue from other buyers. Many sellers offer both: widely available data at a low price and exclusive data, or exclusive periods of access, at a premium.

Signals of quality

Because buyers cannot inspect data fully before purchase, marketplaces and sellers use several signals of quality. Documentation describing how the data were collected, what they cover and what their known limitations are allows buyers to judge fitness for their purpose. Samples and previews let buyers test the data on a small scale. Ratings and reviews from previous buyers provide evidence of reliability. Certification against standards, or audits by independent parties, offers a further check. Each of these signals is costly for a seller of poor data to imitate, which is what gives it value.

For buyers, the main risks are paying for data that are inaccurate, out of date or unlawfully collected, and finding that the data do not cover the population or period they need. Data collected without proper consent may expose buyers to legal liability as well as to reputational damage. Careful buyers ask where data came from, how and when they were collected, what consent was obtained and how the data have been checked.

Data brokers

The largest trade in personal data takes place not on open marketplaces but through **data brokers**, companies that collect information about individuals from many sources and sell it to others. The Federal Trade Commission (2014) found that the brokers it studied held information about nearly every consumer in the United States, combined online and offline sources and sold products for marketing, identity verification and risk assessment. Consumers were largely unaware of these companies and had little ability to see or correct the data held about them. The Commission recommended greater transparency and consumer control.

Zuboff (2019) argued more broadly that some large technology companies have built their business on collecting data about people's behaviour and using it to predict and influence that behaviour, which she called **surveillance capitalism**. Her critique has influenced debate about whether personal data should be traded at all, and on what terms.

Rules for sharing and selling data

Governments have responded with new rules. The European Union's **General Data Protection Regulation** sets conditions for processing personal data, requiring a lawful basis such as consent, limiting use to specified purposes and giving individuals rights to access, correct, delete and move their data (European Union, 2016). These rules affect any marketplace that trades in data about people in the European Union.

The European Union has since adopted rules aimed at encouraging data sharing. The **Data Governance Act** sets up a framework for data intermediaries, requiring them to be neutral and not to use the data they handle for their own purposes, and encourages the reuse of certain public-sector data (European Union, 2022). The **Data Act** gives users of connected products, such as machines and vehicles, rights to access and share the data those products generate (European Union, 2023).

OECD (2019) reviewed policies for improving access to and sharing of data. It concluded that greater sharing could bring large social and economic benefits, but that it required attention to privacy, security, intellectual property and trust, and that a range of approaches, from open data to restricted access under contracts, was needed for different kinds of data.

Alternative models

Not all data sharing takes place through markets.

Open data. Governments and research funders increasingly publish data openly, free of charge, so that anyone can use them. The FAIR principles encourage data to be findable, accessible, interoperable and reusable (Wilkinson et al., 2016). Open data increase the value of public investment in data collection by allowing many users to benefit.

Data trusts and cooperatives. Delacroix and Lawrence (2019) proposed **data trusts**, in which individuals pool rights over their data and appoint trustees to manage them on their behalf, negotiating with those who want to use the data. Such arrangements aim to give individuals collective bargaining power and to ensure that data are used in their interest.

Data collaboratives and research access. Some organisations share data with researchers under controlled conditions, such as secure environments where data can be analysed but not removed. Many statistical agencies provide such access to confidential microdata.

Valuing the individual's contribution

If data about individuals are valuable, should individuals be paid for them? Jones and Tonetti (2020) argued that giving consumers ownership of data about themselves could lead to more beneficial use of data than leaving control with firms, because consumers would weigh privacy against the gains from wider use. Ghosh and Roth (2011) studied how one might design an auction to buy data from individuals who place different values on their privacy, and showed the difficulties that arise when those values are themselves private and related to the data being sold.

Survey research has long faced a version of this question. Respondents give their time and information, often in exchange for a small incentive or none at all, on the understanding that their answers will be kept confidential and used for legitimate research. This model rests on trust and consent rather than on a market price for personal data.

Implications for research organisations

The rise of data marketplaces has several implications for organisations that produce survey and research data.

Collected data have distinct value. Data collected by design, with known methods and with consent for research, differ from data gathered as a by-product of other activities. Their provenance is clear and their meaning well documented, which addresses some of the main barriers to trade identified by Koutroumpis et al. (2020).

Analysis often sells better than raw data. As Bergemann and Bonatti (2019) showed, buyers often value tailored insight more than raw data, which favours providers that combine data with expertise.

Privacy is a condition of trade. Research organisations must ensure that any sharing or sale of data respects the commitments made to respondents and the requirements of data protection law. Aggregated results, synthetic data and controlled access can allow value to be shared without exposing individuals.

Trust is the product. In a market where provenance, quality and lawful collection are hard to verify, an organisation's reputation for transparent methods and responsible data handling is a central part of what it sells.

Conclusion

The commercialisation of knowledge has taken new forms with the growth of data. Data marketplaces aim to connect those who hold data with those who want to use them, providing discovery, quality assurance, pricing, licensing and privacy protection. Yet trading data is harder than trading most goods. Buyers cannot judge value before seeing data, sellers lose control once data are sold, value depends on context, and data about people raise serious questions of privacy. As a result, much trade in data takes place through bilateral agreements, services and analysis rather than open markets. Governments are developing new rules to encourage sharing while protecting individuals, and alternative

models such as open data and data trusts are emerging. For research organisations, the lesson is that well-documented data, collected with consent and handled with care, combined with the expertise to turn data into insight, are well suited to a market in which trust is scarce.

References

Abowd, J. M. (2018). The U.S. Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (p. 2867). ACM. https://doi.org/10.1145/3219819.3226070

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature, 54*(2), 442–492. https://doi.org/10.1257/jel.54.2.442

Agarwal, A., Dahleh, M., & Sarkar, T. (2019). A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation* (pp. 701–726). ACM. https://doi.org/10.1145/3328526.3329589

Arora, A., Fosfuri, A., & Gambardella, A. (2001). *Markets for technology: The economics of innovation and corporate strategy*. MIT Press.

Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In National Bureau of Economic Research (Ed.), *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton University Press.

Bergemann, D., & Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics, 11*, 85–107. https://doi.org/10.1146/annurev-economics-080315-015439

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law, 9*(4), 236–252. https://doi.org/10.1093/idpl/ipz014

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88.

European Union. (2022). Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance (Data Governance Act). *Official Journal of the European Union, L 152*, 1–44.

European Union. (2023). Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data (Data Act). *Official Journal of the European Union, L 2023/2854*.

Federal Trade Commission. (2014). *Data brokers: A call for transparency and accountability*. Federal Trade Commission.

Gans, J. S., & Stern, S. (2010). Is there a market for ideas? *Industrial and Corporate Change, 19*(3), 805–837. https://doi.org/10.1093/icc/dtq023

Ghosh, A., & Roth, A. (2011). Selling privacy at auction. In *Proceedings of the 12th ACM Conference on Electronic Commerce* (pp. 199–208). ACM. https://doi.org/10.1145/1993574.1993605

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review, 110*(9), 2819–2858. https://doi.org/10.1257/aer.20191330

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660. https://doi.org/10.1093/icc/dtaa002

OECD. (2019). *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*. OECD Publishing. https://doi.org/10.1787/276aaca8-en

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, 57*(6), 1701–1777.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. W. W. Norton.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association, 1*(4), 990–1029. https://doi.org/10.1162/154247603322493212

Shapiro, C., & Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business School Press.

Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics*, *54*(4), 208–216. https://doi.org/10.1007/s10272-019-0826-z

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*(5), 557–570. https://doi.org/10.1142/S0218488502001648

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, Article 160018. https://doi.org/10.1038/sdata.2016.18

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.