Policy Brief
Issue n°01/25 | January 2025

U ing da a a a prod c io fac or: polic idea for a r EU da a ra eg

B | Ma |

parties for many purposes at the same time. at creates the potential for economic e ciency gains from the re-use and aggregation of data. Exclusive private data control rights and high transaction costs may stand in the way of realising these gains.

Bertin Martens (bertin.martens@bruegel.org) is a Senior Fellow at Bruegel

C criteria against which existing data market regulations can be evaluated. ese criteria lead to the following recommendations to improve existing EU data regulations:

1. e EU general data protection regulation facilitates re-use of personal data. Machine-readable consent notices and real-time data transfers could reduce high transaction costs that prevent meaningful exercise of informed consent.
2. e European Health Data Space contains an almost-ideal governance regime for health-data re-use and pooling that maximises incentives for data-driven innovation. is regime should be applied to other industrial data-space initiatives.
3. e Data Act facilitates access and re-use of product data but exclusive licensing rights for data holders, monopolistic pricing of third-party data transfers and other anti-competitive measures reduce its impact. Widening its application to services data would make it a truly horizontal data-market regulation.
4. e Digital Markets Act contains several obligations for platforms to grant users access to their own data. Widening access to networked data would facilitate competition between platforms and their users. Mutual instead of unilateral sharing of search-engine data would avoid fragmentation of the search-data pool.
5. e Arti cial Intelligence Act imposes unwarranted and costly restrictions on the re-use of copyright-related content data for AI model training data, reducing the innovative impact of AI. e text and data mining exception in the EU Copyright Directive could be broadened to address these new AI technology challenges.

EU regulations exhibit a tension between exclusive private rights and the realisation of the wider societal value of data. Finding an acceptable balance may involve some redistribution of the e ciency gains between data users and the original data collectors.

R
Martens, B. (2024) 'Using data as a production factor: policy ideas for a new EU data strategy', *P B* 01/2025, Bruegel

1 Introduction

e rapidly growing volume and economic importance of digital data has led European Union policymakers to adopt several data market regulations. Major laws include the general data protection regulation (GDPR, nalised in 2016), the Data Act (2023), the Digital Markets Act (2022), the Arti cial Intelligence Act (2024) and the Regulation on the European Health Data Space (agreed in 2024)[1]. All these regulations seek to open up access to data that is locked up in technical silos, to facilitate the emergence of data markets and to stimulate the development of innovative data-driven services.

e move to make more e cient use of data and leverage its power as a production factor (Beraja and Yuchtman, 2024), similar to labour and capital, is very welcome (as argued by Martens, 2024a). However, the sheer number of EU data laws leads to regulatory fragmentation, increases compliance costs and may result in inconsistencies between regulations (Martens, 2023a). Rules are often precautionary and not as innovation-focused as they could be. at reduces their impact on data markets and data-driven services in the EU. On arti cial intelligence, for example, Draghi (2024) observed that EU restrictions on the storing and processing of data "

2 The efficient use of data as a production factor

Data non-rivalry generates two potential sources of economic bene ts:

- Economies of scope from the re-use of data (Panzar and Willig, 1980; Teece, 1980): once collected, data can be re-used by many parties for many purposes at the same time. For example, the data that Google collects from search queries, data embedded in a bank account or collected by a car, can be re-used for other services and/or by other service providers, to o er complementary and competing services: advertising, payment services, car maintenance services. Re-use by others will not functionally impact the original use but may have an economic impact on the parties that co-generated the original data.
- Economies of scope in data aggregation (Bajari . *a*., 2019; Calzolari . *a*., 2021, Carballa . *a*., 2023): data from many di erent sources can be pooled and aggregated. Data collected by search engines, navigation apps and medical service providers becomes more valuable when aggregated across more users. Pooled data can reveal patterns and deliver service insights that cannot be extracted from fragmented datasets or individual data. For example, navigation services, social-media newsfeeds or search engine recommendations would not be feasible without data aggregation across users.

Teece (1980, 1982) pointed out that the existence of unrealised externalities indicates a failure in markets for complementary inputs required for the production of a service. For example, the holder of car navigation data may not have access to complementary data about hotels and restaurants and is therefore not in a position to o er drivers additional travel services. Collaboration might be tried with a rm that has this additional data but strategic behaviour makes contracting di cult (Schulze . *a*., 2006), especially when there are significant di erences between rms in terms of market power. e data collector may fear that the data will be used against their interests. As a result, data-market failures persist and may require regulatory intervention.

In some cases, markets can overcome obstacles to data re-use and aggregation. For example, Google Maps combines road and navigation data with complementary locational data about businesses and services. Advertising revenue gives it an incentive to invest in harvesting the value of data re-use and aggregation. Consumers are incentivised to contribute their data because they get useful and free navigation services in return for accepting ads. In this case, the market realises at least some of the value of navigation data externalities. But there might be more value from further re-use of navigation data that is not realised yet.

Transaction costs often stand in the way of realising the societal value of data. First, nding partners to share the data with, or to arrange complementary inputs to generate value, may be di cult. Data cannot be exposed in a showroom. e willingness to pay for data may vary between users and service applications. Facilitating exploration of this value may require speci c data-market design (Bergemann and Bonatti, 2018). It is hard to determine the value that data contributes to a data-driven service. Negotiated market outcomes often depend on the market power of the partners. Second, data transfers often require intermediary institutions that de ne data formats and transfer protocols and set the conditions for access and re-use (Martens, 2024b). is can be simple for bilateral data sharing but complex for data aggregation or pooling between many parties.

3 Potential data-market efficiency gains in EU data regulations

3.1 The GDPR

e GDPR is an important 'foundational' data law that regulates markets for personal data collected from natural persons, not from legal entities. It imposes restrictions on the collection of personal data. Firms should ask for the consent of natural persons and should adhere to strict rules on the handling of this data. Personal data cannot be used for purposes other than that for which it was collected. However, the GDPR grants natural persons the right to re-use their personal data for other purposes, or to permit data re-use by other service providers that compete with the original data collector.

at is a pro-competitive and pro-innovation provision: re-use by others increases competition in data-driven services markets in which the original data collector no longer has a data monopoly. ere are no explicit provisions in the GDPR on data aggregation. However, data holders collect data from many persons and are therefore, - *a* . data aggregators. Data holders can combine and pool di erent personal data sources provided it is included in the consent notice.

Unfortunately, use of GDPR rights in practice often runs into high transaction costs. ere is ample empirical evidence that GDPR consent notices are too costly and vague for data subjects to be meaningful (for example Barocas and Nissenbaum, 2009; Cate and Mayer-Schonberger, 2013; Utz . *a*., 2019). Data subjects do not read the many consent notices that pop up during daily web sur ng because it takes too much time and notices are not intelligible. Moreover, data-subject requests for data access and transfers are often met only with considerable delay or in obscure data formats. e GDPR only requires transfers within three months of a request. at delay greatly diminishes the service market value of the data.

All this results in the so-called privacy paradox (Acquisti . *a*., 2016): natural persons attach importance to privacy but in practice do not use privacy protection tools because the costs of doing so are higher than the expected bene ts. Acquisti . *a*. (2016) cited many studies that illustrate how privacy costs and bene ts vary widely according to the setting and the behaviour of data subjects and collectors. It is very di cult for individuals to know how their privacy choices will a ect their welfare. at makes active privacy management very complex.

e GDPR also imposes compliance costs on data service providers. Empirical evidence shows that the GDPR has reduced the supply of digital services in the EU, compared to other regions and to the pre-GDPR period (see Johnson, 2024, for an overview). However, much of that evidence focuses on the supply side. It says little about the impact on consumer welfare or the demand side. Many of the services blocked by the GDPR might have reduced consumer welfare because they use personal data against the interests of the data subject. Others might have increased consumer welfare. How can these two be distinguished? Economists have so far been unable to come up with credible estimates of, or methods to estimate, the economic value of privacy, perhaps because of the wide variation in that value according to circumstances.

Policy recommendations

e GDPR has created the potential for personal-data market e ciency gains through economies of scope in data re-use and aggregation, but policymakers still have some way to go to reduce transaction costs that impede the realisation of these bene ts.

First, onerous transaction costs for consent notices could be substantially reduced by introducing mandatory standardised and machine-readable consent notices. at could generate a more transparent market for consent services, on top of the market for data, and could enable natural persons to delegate that task to specialised service providers that could

handle it in accordance with users’ stated preferences and rms’ stated uses of the data. is would reveal privacy preferences for di erent types of services and consent conditions. A ranking of preferences would be a step towards distinguishing between welfare-augmenting and welfare-reducing personal data services. It would also put pressure on service providers to demonstrate data-sharing bene ts for consumers, as a way to move up the ranking.

Second, making personal data available in real-time through application interfaces (APIs) would greatly reduce transfer transaction costs and make transfers to competing service providers more meaningful in an online digital market setting. Some EU data regulations, such as the Data Act and the Digital Markets Act (see sections 3.3 and 3.4), already include these obligations for data collectors. Nothing prevents the GDPR from doing the same.

3.2 The European Health Data Space (EHDS)

In fact, for one of the most sensitive types of personal data – health data – European data regulators have already gone far beyond the GDPR to generate economies of scope in data re-use and aggregation and in reducing transaction costs. e EHDS is the rst EU data regulation that distinguishes between market failures in data re-use and in data aggregation.
rei.9 efg4(1,(s.-2s9n2T4()e)5.12ets)-1 1s(f)pHgn.1.1 (R fr3an -1.0 -ans)2eansans(1 tot)TJ/T(t s)-3934e Af thd (R frite.9

template in other 'industrial' data-pooling initiatives under the European Strategy for Data (European Commission, 2020). For example, proposals for a Common European Agricultural Data Space (CEADS), designed by farmers' organisations[3], would grant farmers exclusive control rights over farm data, rather than providing shared access rights for data co-producing

Second, the Data Act restricts competition in the re-use of product data. It forbids the re-use of data to design new products that compete with the product manufacturer that initially collected or generated the product data. Data should not be transferred to the platform services of companies designated as 'gatekeepers' (meaning very large, hard-to-avoid platforms) under the EU Digital Markets Act (section 3.4). is prevents a user from transferring data from, for example, smart home appliances to a Google Android or Apple iOS smartphone, or to a Windows computer. It prevents welfare-enhancing network e ects in data re-use and aggregation in digital ecosystems.

Policy recommendations

- e Data Act is the only EU data regulation that allows monopolistic pricing of third-party data transfers and puts anti-competitive restriction on these transfers. at should be abolished. It distorts data markets in favour of product manufacturers and re-introduces the concept of (quasi) ownership rights.
- More fundamentally, the EU should take a clear position against exclusive data ownership rights that have no place in a digital economy, with non-rival data that is co-generated between two or more parties, each with claims to the data. Data should be treated as a co-generated commons to facilitate the e cient use of data as a production factor and to bene t from economies of scope in re-use and aggregation of data.
- e fuzzy category of 'product data' is bound to create confusion. All data resides on a tan-
  -
  gible ph

- Vertically integrated e-commerce platforms that sell products in competition with independent sellers should not use data that is not available to their competitors, to prevent distortions of the competitive level playing eld.
- Gatekeeper search engines should share query, click and view data with other search engines that request access. However, data sharing is not free and can be subject to FRAND pricing.

ese data-sharing obligations are a rst step towards greater data sharing by platforms, beyond the narrow unpaid search and paid advertising data channels that they usually o er to users. e obligations could be extended to encompass a wider set of platform interaction data, beyond rst-party 'own' direct interaction data. For example, consumers usually browse e-commerce platforms and look at several products before deciding on a purchase. Browsing data across products and sellers may provide very useful information for sellers to better understand their competitors. Platforms have this data but often don't share it with sellers. Including networked interactions would give a much better market overview to buyers and sellers on platforms, putting them on par with the quality of the market overview that the platform has. Making this market data sharable between competing platforms would increase competition in otherwise monopolistic platform markets.

Policy recommendations

- Giving business users access to 'their' data implies access to rst-party click-and-view data only. at still leaves the platform operator in a privileged position with more ne-grained market insights. Extending access to second- and third-party network interaction data would enable business users to identify their nearest competing products and sellers and to modify their commercial strategies (Petropoulos , 2023). is could be done using privacy-preserving data access techniques.
- e obligation not to use certain valuable market information is a welfare-reducing lost opportunity to have more e cient markets. It would be better to share that information equally with all relevant market players, rather than not allowing any party to use it.
- Unilateral search-engine data sharing, with data going from the gatekeeper to others, risks fragmenting the search-engine data pool that is important for the e ciency of search. Mutual data sharing between search engines, irrespective of market shares or size, would be a more e cient solution (Martens, 2023a). However, the competitive landscape for search engines may be about to change rapidly under pressure from new AI-driven search tools. is should be taken into account when enforcing this obligation.

to opt-out of free use of these materials under Article 4(3) EU CDSM, thereby reducing the amount of text available for training by 20-25 percent or more. A September 2024 court judgement from Hamburg suggested, however, that commercial AI developers can circumvent the

Finding an acceptable trade-o between these con icting claims to data is a political balancing act that cannot be based on economics only. at trade-o may vary with the evolution of digital technologies and the extent to which they enable data-driven externalities. When new technologies emerge that increase the social value of data because they generate new insights, applications and innovations that were previously not feasible, the pressure mounts to "[illegible]" (Beraja and Yuchtman, 2024) and tone down the exercise of exclusive private rights to enable more extraction of social value from data. Technological progress can also ease the tension when it produces new technologies that make it easier to protect private rights while harvesting the public bene ts. For example, federated learning techniques in AI and machine learning, and various types of privacy sandboxes in personal data protection, leave private data in its own protected setting, while still enabling model training. e balance between private rights and social bene ts is thus a constant political tightrope to be walked, driven by technological progress.

Finding a politically acceptable balance may involve some degree of redistribution of the e ciency gains from data markets between data users and the original data collectors. ere are three options. e rst leaves the gains in the hands of data users – as in the GDPR, the EHDS and the DMA. at maximises incentives for the innovative re-use of data. e second allows data collectors to set a monopolistic price for access to the data, as in the Data Act, the preliminary design of the CEADS and copyright provisions in the AI Act. at reduces competition and innovation incentives, and the e ciency of data as a production factor. An intermediate regime would somewhat soften monopolistic pricing with FRAND conditions, at the cost of substantial administrative intervention and market uncertainty to achieve this. Only the rst option would be fully in line with the EU's oft-stated aim of maximising innovation.

References

Acquisti, A., C. Taylor and L. Wagman (2016) ' e Economics of Privacy', [illegible] 54(2): 442–92

Aplin, T. (2024) ' e Data Act and trade secrets: an experiment in compulsory licensing', in A. Sattler and H. Zech (eds), [illegible], Institut für Recht und Digitalisierung, University of Trier, available at https://doi.org/10.25353/ubtr-04b0-0969-2b7a

Atik, C. and B. Martens (2021) 'Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU', [illegible] 12(3): 370-396

Bajari, P., V. Chernozhukov, A. Hortaçsu and J. Suzuki (2019) ' e impact of big data on rm performance: An empirical investigation', [illegible]

Calzolari, G., A. Cheysson and R. Rovatti (2023) ‘Machine data: market and analytics’, mimeo, available at https://ssrn.com/abstract=4335116

Teece, D.J. (1982) 'Towards an economic theory of the multiproduct rm,' *J a E B a a O a a.* 3(1): 39–63

Utz, C., M. Degeling, S. Fahl, F. Schaub and T. Holz (2019) '(Un)informed Consent: Studying GDPR Consent Notices in the Field,' *P 2019 ACM SIGSAC C C a C a S : 973-990*, available at https://doi.org/10.1145/3319535.3354212