WORKING PAPER ISSUE 21/2023 18 DECEMBER 2023

BERTIN MARTENS

Technical restrictions on access to and re-use of data may result in failures in data markets and data-driven services markets. This paper examines three new EU data regulations (the European Health Data Space, the Data Act and the Digital Markets Act) that vary substantially in mandatory access measures intended to overcome these market failures. It applies three economic criteria, economies of scope in re-use and in aggregation of data, and data supply-side failures, to assess the efficiency of these regulations in overcoming market failures and coherence across regulations. Variations might be justified by particular sectoral market conditions. The European Health Data Space proposal comes close to an ideal data access regime for primary re-use and secondary pooling of health data. The Data Act opens access to data from tangible products only. It strengthens the market power of data holders by giving them quasi-ownership rights over data. It introduces new obstacles to re-use that are likely to minimise its impact. The Digital Markets Act opens access to market data pools collected by very large gatekeeper

Recommended citation:
Martens, B. (2023) "Are new EU data market regulations coherent and efficient?" Working Paper 21/2023, Bruegel

1 Introduction

In 2020, the European Commission published a new European Strategy for Data, comprising a series of regulatory interventions in data markets (European Commission, 2020). This resulted in several horizontal or

data is a by-product of a service that is already paid for. Access to data through regulatory intervention therefore requires careful attention to be paid to the economic implications on the supply side. Similar to the economics of IPR, society requires a balance between exclusive monopolistic rights for investors and access and re-use rights for users. However, a major difference is that creative inventions are produced by one party, the innovator, and used by another party with different interests. Data on the other hand is co-generated between at least two parties

such as IPR and trade secrets, should be protected but cannot be invoked to withhold the data for research purposes (Art. 33 §4) and patients' privacy is protected by means of anonymised or pseudonymised access to the data (Art. 44). However, the identity of medical service providers is not protected. The EHDS imposes purpose limitations with a list of authorised and unauthorised data processing due to the sensitive nature of health data. It allows processing for health research, innovation, policymaking, regulatory and personalised medicine purposes (Art. 34). Any party with a legitimate research purpose can access the data pools. The EHDS only prohibits secondary users from making decisions that are detrimental to the welfare of patients, for example use for the calculation of insurance premia, advertising or marketing activities, or the development of harmful products or services (Art. 35). Findings from secondary use come into the public domain because researchers are required to publish the findings of their research within 18 months.

3 The Data Act: a case of regulatory failure[10]?

Chapters 2 and 3 of the DA target "connected product data" (DA Art. 2 §5 and Art 3), data generated by tangible physical items that can communicate data outside the product. This is a new data category that did not exist before in EU data regulations, and, so far, the DA is the only regulation that makes this distinction. This concept of "product data" emerged first in 2017. A European Commission communication (European Commission, 2017) advocated private ownership rights over "machine data", inspired by Zech (2015), as a means to protect industrial data. The proposed distinction between connected product and other data is rather arbitrary and confusing. Digital data does not float in thin air. All digital data requires a tangible "product" as a physical carrier: a computer to store and process data, and an analogue-digital interface that converts digital data into analogue mechanical and audiovisual signals. These physical carriers may be located in different places, and are owned and operated by different parties. The DA applies only to physical carriers that are directly handled by users.

The DA constitutes an attempt by the EU regulator to overcome control over data exercised by product manufacturers in data-driven services markets. These good intentions are enshrined in DA Art 3 §1, which grants product users direct and free-of-charge access to the product data. This enables economies of scope in the use of data for the purpose of producing competing or complementary data-driven services. Unfortunately, other DA provisions create obstacles for the exercise of access rights, and preserve to a great extent the product manufacturer's monopolistic control over the data.

The original European Commission DA proposal provided access to all data generated by the use of a product. This was subsequently amended to "data of the same quality as is available to the data holder". The text also distinguishes between data stored inside the product or on external servers (DA

[10] This paper only discusses Chapter 2 of the Data Act, on business-to-consumer- and business-to-business data sharing, and Chapter 3 Obligations for data holders to make data available. It uses the version of the Data Act of 7 July 2023, which was approved by the European Parliament on 9 November 2023 (European Parliament, 2023)

Art. 4 §1 and §2). Data transmission from a product to a server is costly. Data holders will limit to data for which they have a private business use. This may exclude data that is valuable to other parties or society at large. Modern cars for example collect thousands of data points manufacturers only collect and use a few hundred of these. It is not clear if the DA would grant car users access to all data available inside a car.

The DA restricts user access and portability to raw data only, ie data without any modification or processing beyond mere conversion of analogue signals into digital formats.[11] This is unfair because it prevents user access to data that was processed as an explicit part of a purchase agreement and that they may have already paid for at the point of sale of the product or subscription to a related service. This provision boils down to a de facto extension of IPR on software to the data outputs of that software.[12] It would be equivalent to, for example, Microsoft retaining an exclusive right over processed data that is generated by Excel worksheets after users put in primary unprocessed data, and charging users when they want to transfer the processed Excel data to a third party. In contrast with the above discussed EHDS is part

third party, they have to pay again for the same data. Users may want to port product data to a third party commercial service provider to obtain competing or complementary services from that party. Although the DA states that users receive the data free of charge, the reality will be that third parties will only want to provide that service if they can charge the user for any additional costs for the acquisition of the relevant data required to provide that service.

Empirical evidence on the impact of third party pricing rules in car maintenance, where manufacturers can charge independent maintenance service providers for access to car maintenance data, shows that it results in an increase of at least 6 per cent in maintenance costs for independent service providers. That distorts competition with service providers affiliated with the manufacturer (Houde, Newberry and Schonenberger, 2019). Applying FRAND pricing equally to all service providers would prevent that distortion. However, it would still result in monopolistic market failure in maintenance services.

The unequal treatment of data generators and the assignment of exclusive rights to product manufacturers and data holders distorts competition and slows down innovation in downstream

locked up in the gatekeeper ecosystem. The underlying problem seems to be that the DA, and the DMA, do not recognise the welfare-enhancing side of network effects and focus only on the monopolistic welfare-reducing side. That brings us to the DMA itself.

The DA also mentions trade secrets in digital data.[17] Trade secrets should not prevent access to data, other than in exceptional circumstances when the product manufacturer could suffer extreme harm. However, they shall be disclosed only where the data holder and the user take measures to preserve their confidentiality, in particular regarding third parties. Moreover, it is up to the trade secret holder to identify the data that he considers to amount to trade secret. It is unclear what related data trade secrets mean in a digital context. The EU Trade Secrets Directive (Directive (EU) 2016/943) defines three conditions for the existence of trade secrets: (a) the information is not known either by the public at large or by the experts of the sector; (b) the information has commercial value; and (c) the claimant has taken steps to keep the information secret. Following these conditions, the trade secret status of market information may vary according to the level of data aggregation. For example, data about a single sale is not a secret for the seller because the buyer has the same information. Aggregated sales data, the turnover of a seller, might constitute a trade secret for the seller, though the platform has this information too. The seller's market share on a particular platform is known to the platform operator only and cannot be a trade secret for the seller and not for the platform. Data-related trade secrets will need to be defined better.[18]

In contrast to the EHDS, the DA focuses on primary data access and portability only, ie the benefits from economies of scope in the use of data. It does not seek to generate economies of scale and scope in data aggregation and secondary use in data pooling. The European Commission's European Strategy for Data (European Commission, 2020) states that sectoral data pools will be the subject of separate policy initiatives. Some of these have already been launched, for example in agriculture and mobility data,[19] though there are as yet no details on data governance proposals for these pools.

4 Access to market data pools: the Digital Markets Act

The DMA is first and foremost a competition policy instrument that seeks to rein in the anti-competitive behaviour of very large platforms that have become dominant because of gatekeeper network effects: more users make a platform more interesting to other users and therefore attract more users. More users also leave more data traces that enable a platform to improve the quality of user matching services which, again, attracts more users. Network effects crowd out competitors and 'tip' market towards a single dominant platform. Users then suffer from the monopolistic impact of network effects: reduced choice and increased prices exceed user benefits from network effects. The DMA

[17] Notably in DA Recital 31 and Art 4.3 and 5.8.
[18] See for example Apella (2023).
[19] See https://digital-strategy.ec.europa.eu/en/library/common-european-data-spaces-agriculture-and-mobility.

imposes obligations on gatekeepers to restrict their monopolistic behaviour, weaken network effects and stimulate competition, including through three data sharing obligations.

First, gatekeepers should give business users and end users (consumers) access to the "data generated by their activities on the platform" (DMA Art. 6 §10). That enables economies of scope in the re-use of data. This obligation is an extension from personal data of end users to business user data of GDPR rights, and from delayed to real-time access to personal data.

Second, the DMA seeks to level the information playing field between a vertically integrated gatekeeper and its business users. Gatekeepers are not allowed to make privileged use of their ma data to compete with business users on their platform (Art. 6 §2). They can only use that data when they have also made it available to business users.

Third, gatekeeper search engines – in practice, Google Search – should share "query, ranking and data" with competing search engines (Art. 6 §11). Search engines collect data on user queries and clicks on webpage rankings that the search engine delivers in response to a query. Search engines crawl billions of webpages and select and rank these to respond to queries. By observing user click on the proposed page rankings, they learn how to better respond. More frequently clicked pages m up the ranking. Since most queries are rare, climbing the learning curve may be slow. More users usin the search engine improves data collection and delivers more efficient responses, even to rare queries. Better responses, in turn, attract even more users. User data driven network effects explain why a single search engine became dominant.

The first two obligations suffer from lack of clarity about the extent of data sharing. Data generated by their activities on the platform implies access to interaction data with other users, and data in the form of platform responses to user queries. For example, in an e-commerce platform, user activities necessarily entail interactions with products and services offered by sellers. When gatekeepers should make market data available to competing business users, what level of fine-grained market data should be made available to whom and under what conditions? To restore a market information level playing field, it should clearly go beyond business users' interaction data in the platform. Martens et al. (2023) suggested that second-degree network interaction data should be sufficient to enable business users to position themselves more efficiently in a platform marketplace and compete with vertically integrated sellers. The third obligation for gatekeeper search engines to share query and click data with competitors is very far-reaching and comprises the search engine's entire aggregated dataset, including user query inputs, search engine responses and clicks on these responses. It makes the full search engine data available to competitors.

Access to user interaction data goes beyond enabling users to benefit from economies of scope in the re-use of data. Network interaction data has a data pooling dimension across many users. Access to this data gives users access to economies of scale and scope in data aggregation. The DMA thus fo gatekeeper platforms to share the benefits from network effects with competitors, thereby levelling the data playing field between competitors. By analogy to the terms of data sharing provisions in the

5 Discussion and conclusions

All three EU data regulations discussed in this paper facilitate access to data held by companies. While the EHDS puts almost no conditions on access, the DA imposes very stringent conditions, including payment of a monopolistically-priced license fee to the data holder, who becomes a quasi-owner of the data in case of third-party portability, and the prohibition on use of the data to compete with the data holder. The DMA puts no conditions on access to own platform data by natural persons and business users, but attaches quasi-exclusive ownership rights, somewhat attenuated by "fair" pricing conditions, to search engine data.

Only the EHDS has explicit provisions for data pooling. There are none in the DA. The European Strategy for Data announced that the creation of and access to sectoral data pools will be regulated in separate and still-to-be-announced policy instruments, outside the DA. Gatekeeper platforms targeted by the DMA could be considered as market data pools however. In that sense, the DMA regulates access to privately created and very large market data pools. It restricts that access to narrowly de ned users' "own" data, not to the full pool of user interaction data. Only in the case of marketplace and search engine data are platforms under the obligation to share a much broader, but not very clearly de ned, interaction dataset.

All three regulations remain vague, and sometimes inconsistent, about access to processed user data. The EHDS does not distinguish between raw and processed data; it grants access to all personal health data. In the DMA, access to marketplace and search engine data also includes access to processed data. It fudges the question of whether users' access to their own data includes processed user interaction data on the platform. The DA opens access to data as available to the product manufacturer or data holder, but then backtracks and limits access to raw or "not substantially" processed data. The EU GDPR was the first data regulation to restrict personal data access rights to data "contributed by" the data subject. This restriction becomes hard to maintain in the DA when processed data is part of the services related to a product that the user already paid for at the point of sale or subscription to a service: why should users not be granted access rights in that case?

All three regulations frequently assert the primacy of personal data protection rules under the GDPR. However, the EHDS and DA also refer to the need to protect trade secrets. Only the DMA does not refer to that subject, at least not in the context of mandatory data sharing. It is unclear how to de ne trade secrets in data when data is generated between two or more parties.

Returning to our initial question, would one EU data regulation instrument be enough, or do we need many regulations to cover the variety of circumstances in different sectors? The comparison of the three data regulations shows that the EHDS is an example of a near-ideal data regulation that ticks almost all the boxes for maximum economies of scope in primary and secondary economies of scale and scope in data pooling. From the point of view of overcoming regulatory market failures, it would have been a better sectoral regulatory template than the DA. Applying the EHDS template for primary use would have resulted in dropping the superfluous and confusing concept of product

References

Kerber, W. (2022) 'Governance of IoT Data: Why the EU Data Act will not fulfill its Objectives', GRUR International 72(2):120–135

Ledyard, J. O. (2008) 'Market failure', in The New Palgrave Dictionary of Economics, 2nd edition

Martens, B. (2023) 'Pro- and anti-competitive provisions in the proposed European Union Data Act', TILEC Discussion Paper 2023-03, Tilburg University

Martens, B., G. Parker, G. Petropoulos and M. Van Alstyne (2021) 'Towards Efficient Information Sharing in Network Markets', TILEC Discussion Paper 2021-014, Tilburg University

Panzar, J. and R. Willig (1981) 'Economies of Scope', American Economic Review 71(2): 268–272

Perzanowski, A. and J. Schultz (2016) The End of Ownership: Personal Property in the Digital Economy, MIT Press

Posner, R. (1978) 'The Chicago School of Antitrust Analysis', University of Pennsylvania Law Review vol. 127: 925–948

Teece, D. (1980) 'Economies of scope and the scope of the enterprise', Journal of Economic Behavior & Organization 1(3): 223–247

Zech, H. (2015) 'Information as property', JIPITEC 6(3): 192–197

Bruegel, Rue de la Charité 33, B-1210 Brussels
(+32) 2 227 4210
info@bruegel.org
www.bruegel.org