Martens, Bertin

Working Paper

Are new EU data market regulations coherent and efficient?

Bruegel Working Paper, No. 21/2023

Provided in Cooperation with:
Bruegel, Brussels

Suggested Citation: Martens, Bertin (2023) : Are new EU data market regulations coherent and efficient?, Bruegel Working Paper, No. 21/2023, Bruegel, Brussels

This Version is available at:
https://hdl.handle.net/10419/294884

WORKING PAPER | ISSUE 21/2023 | 18 DECEMBER 2023

ARE NEW EU DATA MARKET REGULATIONS COHERENT AND EFFICIENT?

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 platforms. Some access provisions are vaguely defined. Others facilitate access to data pools but may risk unwinding the benefits of data-driven network effects. There is scope for significant improvement in these data regulations.

Bertin Martens (bertin.martens@bruegel.org) is a Visiting Fellow at Bruegel and a research fellow at the Tilburg Law and Economics Centre, Tilburg University

bruegel

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 cross-sectoral data regulations, including the Data Governance Act (Regulation (EU) 2022/868), the Data Act[1] and sector-specific regulations, such as the European Health Data Space (European Commission, 2022b)[2], an announced policy proposal on vehicle data[3] and several sectoral data-pooling initiatives in agriculture, transport, energy, etc. Moreover, the Digital Markets Act (Regulation (EU) 2022/1925), a competition policy tool that targets very large digital 'gatekeeper' platforms, also includes data-access obligations. While it is too early to assess their actual economic impact, this paper compares and assesses the potential economic impact of three of these EU data market regulations: the European Health Data Space (EHDS), the Data Act (DA) and the Digital Markets Act (DMA).

It Is easy for stakeholders to get lost in this flurry of sometimes partly overlapping data regulations that contain a wide variety of rules about who can access what data under which conditions. This raises a fundamental regulatory design question: should data regulations be tailor-made to fit the particular circumstances of each sector or issue, or would it be better to have a single horizontal regulation with similar rules for all sectors and domains? One could reformulate this question and ask if data market failures follow a general pattern across sectors, or if there are specific data market failures in some sectors or domains that merit a specific regulatory solution. To answer this, criteria are needed to assess problems in data markets and to evaluate the design of data regulations that seek to overcome these problems.

We apply the well-known economic criterion of market failure (Ledyard, 2008) to address these questions. Regulatory intervention is justified when a market fails to operate in a socially optimal way, ie when it does not deliver the potential social welfare for society as a whole that it could potentially deliver, often because private operators have no incentive to behave in a socially optimal way. The market failures approach is recommended by the European Commission's own Better Regulation Guidelines and Toolbox (European Commission, 2021, 2023). Our assessment revolves around three economic characteristics of data that are at the source of most data market failures: economies of scope in the re-use of data, economies of scale and scope in data aggregation, and market incentives to invest in data collection. We examine the measures proposed in the EHDS, the DA and the DMA to overcome these market failures.

[1] European Commission (2022a). The discussion in this paper is based on the post-trialogue version voted in the European Parliament on 9 November 2023 (European Parliament, 2023).
[2] Still being discussed by the European Parliament and Council of the EU at the time of writing.
[3] See European Commission call for evidence on 'Access to vehicle data, functions and resources', 23 March 2022, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13180-Access-to-vehicle-data-functions-and-resources_en. No policy initiative has been published at the date of writing.

Once data is collected, non-rivalry enables unlimited re-use for many purposes, without any functional impact on the original use for which the data was collected. Re-use is beneficial for society, though not necessarily for the data collector. Data collection however is often rival because it requires access to the physical device used by an agent at the moment of collection. Non-rivalry can generate economies of scope in the re-use of data[4], by the data holder and/or by a third-party. The data holder may block re-use, especially when that third party is a potential competitor to the data holder. Blocking re-use is a data market failure and may also result in a monopolistic market failure in downstream services markets[5].

A unique characteristic of data is the potential for economies of scale and scope in data aggregation or pooling[6]. These emerge when more valuable insights can be extracted from pooled data compared to fragmented datasets, when the combined social value of data exceeds their private value to individual data holders. Private entrepreneurs may create data pools by sharing the benefits from economies of scale and scope in aggregation with users who contribute their data to a pool. For example, an e-commerce platform pools data from buyers and sellers. Here, the gains are realised through network effects that benefit all users. The platform operator solves a collective-action problem among data pool users. However, in many cases, private data markets underperform and prevent the full realisation of the social value of the data. Private incentives for pooling are often weak[7] because potential participants fear losing control over their data or disagree with the distribution of benefits from the data pool. That may justify regulatory intervention to facilitate pooling and overcome private disincentives to the production of the full social value of the data.

The third market failure criterion revolves around excludability of data. Without excludability, private investment in data collection is risky because it invites free-riding by others. Excludability of non-rival products is often achieved by means of exclusive property rights for a single party, for example in intellectual property rights (IPR). In the absence of legal property rights over data[8], investors may apply Technical Protection Measures (TPMs) to ensure exclusive control over data and recuperate investment costs in data collection, storage and processing. Mandatory data sharing at zero cost may erode the incentive to invest in data collection and result in a negative data supply response, unless

[4] Economies of scope in re-use were defined long before the ascent of the digital data economy. See Panzar and Willig (1981); Teece (1980).

[5] According to the Chicago critique (Posner, 1978), this monopolistic market failure occurs only if aftermarket conditions are not transparent at the point of sale in the primary market. The data holder can use the data to set up a new service through vertical integration in a service market. Using the data for another purpose may require acquisition of complementary inputs to combine with the data. Failure to access these inputs blocks market entry for the data holder (Teece, 1981). Co-production may also fail because co-producers may not reach an efficient resource sharing agreement (Schultze *et al*, 2005).

[6] For a more detailed discussion of this concept, see for example Calzolari *et al* (2022), Carballa *et al* (2021) and Bajari (2018).

[7] Calzolari *et al* (2022) concluded that participants may end up in a non-cooperative Nash equilibrium that is not Pareto-efficient.

[8] One might argue that the *sui generis* right under the EU Database Directive (Directive 96/9 (EC)) is an exception. However, the *sui-generis* right applies to databases, a structured set of data, not to individual data.

data is a by-product of a service that is already paid for. Opening 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: a data service provider and a user. Both parties may claim rights over the data but may have conflicting interests. That in itself requires a more open approach.

All three EU data regulations discussed in this paper aim to overcome these data market failures by granting conditional access rights to the parties that co-generated the data, or to third parties. But they do so under very different conditions. Data regulations span a policy spectrum from very closed, with strong control rights for private data investors and holders, to very open, with wide-ranging access rights for other co-generators and third parties and thus facilitating the realisation of the social value of data. This paper (a) describes variations in the balance between private and social rights to data across three EU data regulations; and (b) explores if there is room to improve that balance and overcome data market failures more efficiently, ie generating more social welfare from private data. The key criterion is: can societal benefits from data be increased without undermining private incentives to invest in data collection?

Section 2 starts with the European Health Data Space (EHDS), a regulatory proposal still under discussion at the time of writing. The reason for bringing this sector-specific data regulation proposal to the forefront is that it ticks nearly all the data market failure boxes and solutions. It could be considered as 'best practice' in data regulation. At the other extreme of the spectrum stands the Data Act (DA), discussed in section 3. The DA is meant to be a horizontal template for 'product' data across all sectors. But it suffers from excessive protection of data holders, giving them quasi-ownership rights to the data, at the expense of product users. It contains a mix of pro- and anti-competitive provisions, some of which may even worsen market distortions. Section 4 turns to the Digital Markets Act (DMA). This is first and foremost a competition policy tool to overcome monopolistic market failures in the services offered by very large digital gatekeeper platforms, some of which may be caused by exclusive platform control over user data. The DMA facilitates access for persons and business to their 'own' data. That may be too restrictive when data is co-generated in interactions between several parties. This paper argues that widening access to interaction data is important to level the playing field in downstream data-driven services markets. Section 5 discusses the findings.

2 Best practice in data regulation: the European Health Data Space

The EHDS ticks all the boxes in the above-described economic criteria for optimal data regulation. It facilitates the 'primary' re-use of personal health data (EHDS Article 3) and establishes the conditions for 'secondary' health data aggregation in national and EU-wide data pools (Art. 33), managed by public health authorities (Arts. 10 and 36). It puts no restrictions on primary re-use of health data at the initiative of the patient, and few restrictions on secondary re-use of aggregated health data[9]. There are no charges for primary and secondary re-use other than the marginal cost of accessing the data (Art. 42). Charging monopolistic prices is not allowed. This pricing rule implies that all innovation benefits accrue fully to the innovator. Data suppliers cannot claim a share of these benefits.

All human health data is, by definition, personal data. The right to personal data portability, at the initiative of the data subject and at zero cost, is already foreseen in Art 20 of the EU General Data Protection Regulation (GDPR, Regulation (EU) 2016/679). However, in practice, the exercise of that right encounters many hurdles because the GDPR remains vague on the operational aspects of portability. The EHDS fills that gap. It defines six priority categories of health data that should be available for portability: patient summaries, e-Prescriptions, medical images, medical reports, laboratory results and patient discharge reports (Art. 5). It also sets rules to operationalise real-time data portability between Electronic Health Records (EHR) systems operated by medical services supplies in the EU (Art. 6), and defines the obligations of EHR suppliers to ensure this interoperability (Arts. 17-27). The GDPR excludes portability of processed data that is assumed to be under the exclusive control of the data processor. The EHDS however extends portability to processed health data. The six categories include processed personal health data, for example in the form of medical diagnosis and treatment recommendations.

Primary health data portability is unlikely to induce negative health data supply effects. The six standardised datasets are a by-product of medical services delivery. The cost of collecting that information is already borne by the cost of medical services, paid for by patients and medical insurance providers. Doctors and hospitals will not dispense fewer medical services because the data is re-used elsewhere. There may be additional investment costs for health service providers for setting up the re-use infrastructure, some of it possibly borne by public health authorities. Since these six standard datasets are mandatory, service providers cannot avoid these costs. The EHDS does not contain incentives however to supply medical data beyond the standardised dataset. Additional incentives may be required for that purpose. For instance, access to digitised surgery data may require substantial investments by hospitals.

For secondary use of national health data pools for research purposes, public and private healthcare providers are obliged to make fifteen categories of data available, including the six categories of EHR data, and extending into other areas such as genetic data (Art. 33 §1). Prior private rights to these data,

[9] EHDS Art. 35 prohibits uses that would harm persons or societies or exclude persons from health insurance, and use for advertising purposes.

such as IPR and trade secrets, should be protected but cannot be invoked to withhold the data for research purposes (Art. 33 §4). 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 a 2017 European Commission communication (European Commission, 2017) that 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 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 monopolistic control 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 re-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 final trialogue 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 retrieval to data for which they have a private business use. This may exclude data that has value to other parties or society at large. Modern cars for example collect thousands of data points, but car manufacturers only collect and see business value in 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 *"substantial modification"* or processing[11], beyond mere conversion of analogue signals into digital formats. 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. The contrast with the above-discussed EHDS is particularly salient here. Imagine that producers of x-ray equipment or laboratory analysis equipment would retain an exclusive property right over the processed data outputs of their machines.

Apart from legal recognition of manufacturers' exclusive rights to the processed data, the DA also endorses quasi-ownership rights to unprocessed primary data[13]. This is reflected in the provision that data holders or product manufacturers can charge third parties, when they are businesses, a price for data ported to them (DA Art. 9). That price can be based on the fixed costs as well as variable costs for data collection, storage, processing and transmission. Moreover, they can charge a monopolistic price with a mark-up margin. Only SMEs escape from monopolistic pricing (Art. 9 §4). This boils down to a licensing fee for data access, similar to a licensing fee for IPR holders. The DA tries to soften the blow by recommending a *"reasonable"* profit margin and Fair, Reasonable and Non-Discriminatory (FRAND) pricing (Art. 9 §1), a controversial topic in standard essential patents, where FRAND pricing was first applied[14]. While it may not be clear to economists how to calculate a FRAND price with a reasonable profit margin, the DA instructs the European Commission to set up guidelines for that calculation (Art. 9 §5).

Far from FRAND, this pricing rule is unfair because users pay twice – and product manufacturers get paid twice – for the data that they co-generated. At the point of sale, rental or subscription of the product, users pay the product manufacturer for the hardware and software that generates, processes and transmits primary and processed data, and possibly for additional processed data services through subscriptions. When users subsequently want to port this primary and processed data to a

[11] DA Recital 15. This recital is very explicit about the extension of IPR rights on proprietary software and algorithms to the data produced by that algorithm or software.
[12] The creeping extension of copyright on algorithms to data produced by algorithms is a wider phenomenon in the digital economy that has been discussed extensively by Perzanowski and Schultz (2016).
[13] For a more detailed discussion of quasi-ownership rights in the DA, see Kerber (2023) and Martens (2023).
[14] For a discussion, see for example Habich (2022).

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 produce 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 percent in maintenance costs for independent service providers. That distorts competition with service providers affiliated with the manufacturer (Hoegaerts 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 co-generators and the assignment of exclusive rights to product manufacturers and data holders distorts competition and slows down innovation in downstream markets for data-driven services. This constitutes a regulatory failure. We attribute this to the ghost of the European Commission (2017) Communication on data ownership rights that is still hovering over the DA, not only with the introduction of the *"product data"* category that comes close to *"machine data"*, but also with the assignment of IPR-like quasi-ownership control and pricing rights to data that over-protect product manufacturers and/or data holders at the expense of users.

Moreover, the DA legalises further distortions in downstream data-driven product and services markets by prohibiting the use of data for competitive purposes, to compete with products and services produced by the manufacturers and/or data holders (DA Art. 4 §10).

The DA[15] prohibits data transfer from data holders to third-party platforms and services that have been designated as gatekeepers under the DMA, even when requested by the product user. However, it leaves open the possibility that users transfer data directly from their device to a gatekeeper. The data architecture of the product therefore matters. If data is available on the product, users can freely choose a third-party destination, including gatekeepers. If data is stored on a cloud server operated by the data holder, transfers to gatekeepers are prohibited[16]. For example, users of smart home appliances that store data in the cloud cannot transfer the data to their Apple or Android smartphones. That prohibition may destroy potential consumer value from interoperable components of data ecosystems. Regulators have tried to justify this prohibition with the argument that monopolistic gatekeeper platforms should not be given access to even more data than they already have; it would only strengthen their market positions. The counter argument is that the DMA already imposes obligations on gatekeepers to provide users access to and portability of gatekeeper data. Data is not

[15] DA Art. 5 §3, in particular, Art. 5 §3c.
[16] This prohibition could be circumvented when users claim that product data is personal data. In that case, third-party data portability rights under Art. 20 GDPR would apply, allowing users to freely choose a third-party destination of their choice.

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 a trade secret. It is unclear what data-related 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 that 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, nor for that 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 re-use of data. It does not seek to generate economies of scale and scope in data aggregation, or 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 reign in the anti-competitive behaviour of very large platforms that have become dominant gatekeepers because of 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' a market towards a single dominant platform. Users then suffer from the monopolistic impact of network effects: reduced choice and increased prices may exceed user benefits from network effects. The DMA

[17] Notably in DA Recital 31 and in Art. 4 §6.
[18] See for example Aplin *et al* (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) real-time 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 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 market data to compete with business users on their platform (Art. 6 §2). They can only use this data when they have also made it available to business users.

Third, gatekeeper search engines – in practice, Google Search – should share *"query, ranking and click 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 clicks on the proposed page rankings, they learn how to better respond. More frequently clicked pages move up the ranking. Since most queries are rare, climbing the learning curve may be slow. More users using the search engine improves data collection and delivers more efficient responses, even to rare queries. Better responses, in turn, attract even more users. User-driven and 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. User data generated by their activities on the platform implies access to interaction data with other users, and to processed 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, this should clearly go beyond business users 'own' 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 clicks data with competitors is very far-reaching and comprises the search engine's entire aggregated dataset, including user query inputs, search engine responses and users' clicks on these responses. It makes the full search engine data pool 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 forces 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

EHDS, this goes beyond *"primary re-use"* of own data and would be equivalent to *"secondary re-use"* of pooled data.

Regulators should be careful however with sharing of pooled data to avoid weakening network effects, because doing so may be welfare-reducing for users (Martens, 2023). To the extent that Google Search's market share declines when it shares data with competitors and more competing search engines enter the market, the quality of Google Search will also decline because it collects less user data and the size of its data pool will diminish. As a result, competitors will learn less from access to Google's data, especially in the long tail of rare queries. The quality of competitor search services will not exceed the declining quality of Google Search. Consequently, the efficiency of all search engines will decline, and so will user welfare, with the weakening of data pooling and network effects. This problem could be overcome easily by replacing asymmetric data sharing from gatekeepers to competitor search engines with symmetric data sharing between all search engines, irrespective of market share. That would preserve the complete search engine data pool and thus economies of scale and scope in search data aggregation. Unfortunately, symmetric sharing is not foreseen in the DMA.

Platform data-sharing obligations are unlikely to have a negative impact on data collection because data is the by-product of platform services that are already paid for in their business models. However, the search engine case shows that the design of data-sharing rules may be important in this respect.

Moreover, like the DA, the DMA also contains FRAND data pricing for search engine data (Art. 6 §11). The reason for this rule is not explained but one might presume that this is meant as a – superfluous – incentive to keep collecting data. Data collection is already incentivised by the advertising revenue that search engines generate. It gives the search data operator an exclusive quasi-licensing right on search data. It is hard to define what FRAND means in this market. Data on rare queries is more valuable than data on common queries. Smaller search engines would have higher willingness to pay for a larger dataset but less capacity to pay because of lower advertising revenue – assuming that this remains the standard search engine business model. The FRAND condition would not allow price discrimination between search engines. As discussed in the DA section, data pricing reduces data sharing and thus the welfare benefits from economies of scope in the re-use of data.

Note that the DMA does not mention trade secrets as a possible limiting factor on gatekeepers' data sharing obligations. Trade secrets are only mentioned in the context of the regulator's reporting on gatekeepers' compliance with DMA obligations.

5 Discussion and conclusions

All three EU data regulations discussed in this paper facilitate access to and re-use of 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 for 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 defined 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 wider, but not very clearly defined, 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 the same 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 raw 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 has 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 define trade secrets in data when data is co-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 nearly-ideal data regulation that ticks almost all the boxes for maximum economies of scope in primary re-use and secondary economies of scale and scope in data pooling. From the point of view of overcoming data re-use market failures, it would have been a better cross-sectoral regulatory template than the DA. Applying the EHDS template for primary re-use would have resulted in dropping the superfluous and confusing concept of product

data, allowing access to processed data that users have paid for, avoiding users having to pay twice for data in the case of third-party portability, and dropping restrictions on data use for competitive purposes. Similarly, the EHDS template for primary re-use would have been a better recipe for users' access to their 'own' platform data under the DMA. It would unequivocally widen these access rights to processed direct interaction data.

The EHDS template for secondary access to data pools could also have been applied in the DMA, to give business users access to marketplace and search engine data pools. As pointed out, care should be taken to preserve the integrity of data pools in order not to weaken economies of scale and scope in data aggregation. The DMA's asymmetric data sharing obligations for search engines risk promoting competition at the expense of fragmenting that pool and thereby reducing user welfare. Symmetric data sharing, as in the EHDS, would be the preferred solution.

However, the EHDS and the DA show that it is not enough to just define access rights. They cannot be implemented without overcoming the technical obstacles to data access and portability. That requires technical standards that are likely to be specific by sector and/or domain. The EHDS and the DA pay attention to standard-setting procedures. The EHDS defines the medical dataset that should be made available. The DA covers many sectors and includes general provisions that leave room for initiatives to set data standards in various domains. The DMA still has to define standards for data sharing within and between platforms. That will require more regulatory work and instruments.

We conclude from this comparison that there is significant scope to improve data-access provisions in the Data Act, and to some extent in the DMA.

References

Aplin, T., A. Radauer, M.A. Bader and N. Searle (2023) 'The Role of EU Trade Secrets Law in the Data Economy: An Empirical Analysis', *IIC International Review of Intellectual Property and Competition Law* 54: 826–858

Calzolari, G., A. Cheysson and R. Rovatti (2022) 'Machine Data: market and analytics', mimeo, European University Institute

Carballa-Smichowski, B., N. Duch-Brown, S. Höcük, P. Kumar, B. Martens, J. Mulder and P. Prüfer (2022) 'Economies of scope in data aggregation with a case study in health data', mimeo, available at https://www.researchgate.net/publication/365276562_Economies_of_scope_in_data_aggregation_evidence_from_health_data

Bajari, P., V. Chernozhukov, A. Hortaçu and J. Suzuki (2018) 'The impact of big data on firm performance, an empirical investigation', *NBER Working Paper* 24334, National Bureau of Economic Research

European Commission (2017) 'Building a European Data Economy', COM(2017)9

European Commission (2020) 'A European strategy for data', COM/2020/66 final

European Commission (2021) 'Better Regulation Guidelines', SWD(2021) 305 final, available at https://commission.europa.eu/system/files/2021-11/swd2021_305_en.pdf

European Commission (2022a) 'Proposal for a Regulation of the European Parliament and Council on harmonised rules on fair access to and use of data (Data Act)', COM(2022)68

European Commission (2022b) 'Proposal for a regulation of the European Parliament and the Council on the European Health Data Space', COM/2022/197 final

European Commission (2023) 'Better Regulation Toolbox', July 2023 edition, available at https://commission.europa.eu/system/files/2023-09/BR%20toolbox%20-%20Jul%202023%20-%20FINAL.pdf

European Parliament (2023) 'Legislative resolution of 9 November 2023 on the proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)', P9-TA(2023)0385

Habich, E. (2022) 'FRAND Access to Data: Perspectives from the FRAND Licensing of Standard Essential Patents for the Data Act Proposal and the Digital Markets Act', mimeo, University of Zurich

Hoegaerts L. and B. Schonenberger, (2019) *The automotive digital transformation and the economic impacts of existing data access models*, Technical Report, FIA Region I

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.D. 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