# Data in digital markets

Prepared for
European Commission

30 September 2018

www.oxera.com

# 1      Introduction

*This paper is a contribution by Oxera to the European Commission's one-day conference entitled 'Shaping competition policy in the era of digitisation', to be held on 17 January 2019. It is one of three Oxera contributions, the other two covering the topics of algorithms and market power in digital platforms. Oxera is a European economics consultancy specialising in competition and regulation, and actively participating in key policy debates on digital markets.*

Data is everywhere in the online world, be it as a means of exchange for services, an input for customisation or machine learning, or as variables that feed into matching algorithms. Data has gained increasing attention from academics and policymakers. Instances of data misuse—such as the Cambridge Analytica case in early 2018—brought to the public's attention the scale of the data collected about the everyday lives of consumers and how it can be misused.[1]

An increasing number of businesses rely on data to add value by matching consumers and suppliers, generating revenues through targeted advertising, or by automating and refining processes through machine learning. This combination of data and technology has enabled innovation in goods and services while providing increased choice and lower prices for consumers.

Many different types of data feed into a broad range of purposes, hence requiring a differentiated approach to capture the various effects of data and to assess the impact of potential interventions. At the same time, the digital world is expanding and linking devices such as cars, watches and thermostats to the Internet. The Internet of Things (IoT) further contributes to the rapid growth of data collected online.

The importance of this very same data has created concerns about data concentration and, in the case of personal data, privacy. A large share of data residing with a few firms could represent a barrier to entry, limiting competition. Meanwhile, in the context of personal data, consumers do not always know or understand where or how their data is being collected or used, and firms might fail to provide consumers with adequate transparency and control over this.

Over the course of this paper, we will briefly summarise the impact of data on online markets, set out a framework for distinguishing between different types of data, and discuss implications for market outcomes.[2]

---

[1] See Information Commissioner's Office (2018), 'ICO statement: investigation into data analytics for political purposes', 2 May.
[2] Parts of this article are based on Oxera's report on consumer data in online markets, Oxera (2018), 'Consumer data in online markets', 5 June. The report was commissioned by Which? as part of a project regarding the collection and use of consumer data. See Which? (2018), 'Control, Alt or Delete? The Future of Consumer Data', policy report, June.

# 2 Consumer data and online services

Consumer data has had a strong impact on services provided online. It has strongly reduced the search cost for consumers seeking existing services, while it has also facilitated the provision of new services that rely on data as an input.

## 2.1 The impact of data on existing services

Access to data has made it easier to 'match' consumers with products or services—indeed, most online platforms and services rely on consumer data in their matching processes.[3] There are often at least two aspects to this process—past data allows firms to detect patterns of consumer behaviour, while current personal data may allow them to customise their product or service.

Data in and of itself is less likely to change outcomes in markets but combined with advances in algorithmic technology it can lead to significant changes.

The rise of ad-funded business models that allow the offering of services to consumers at a lower price or a price of zero, such as Facebook, is also partly the result of greater access to consumer data.[4] These platforms use consumer characteristics to match consumers with suitable adverts, generating their revenues from advertisers (on the other side of the market).

## 2.2 The impact of data on new services

Access to data also has an impact on new services. For example, online data can render less important personal interactions to gather information such as between consumers and personal trainers or doctors. New firms can use data to offer similar services without having to be in the same location and at a lower cost. This is the case for fitness apps and activity-tracking devices, as well as health chat bots.[5]

Artificial intelligence draws on data as an input and drives innovation in many forms. Machine learning algorithms process large datasets, thereby automating complex processes such as voice recognition or spatial perception. These processes form part of innovations such as digital personal assistants and autonomous driving.

---

[3] Oxera (2015), 'A fair share? The economics of the sharing economy', *Agenda*, December.
[4] Facebook Business (2018), 'Choose your audience', https://en-gb.facebook.com/business/products/ads/ad-targeting, accessed 28 September.
[5] For example, see https://www.babylonhealth.com/about, accessed 28 September.

# 3    Economic characteristics of data

Use of data can affect market outcomes in terms of competition and privacy, depending on the economic characteristics of that data. Only personal data is relevant for privacy outcomes; however, anonymised data is structurally similar as it excludes information on the individual(s) involved in the data-generating process (such as using a phone or driving a car).[6]

The extent to which different firms are able to access similar data is important for competition. In principle, data as a digital good is non-rivalrous, i.e. possession and use by one firm does not preclude other firms from accessing it. In practice, whether different firms can access similar data depends on two key factors.[7]

- **The cost of acquiring data**, which in turn depends on how the data is collected. Broadly, this can happen in three ways: first, people may actively choose to provide their data (e.g. payment details); second, the data is observed from their behaviour (e.g. browsing history or driving behaviour); or third, the data is inferred through analysis of previously acquired data (e.g. personality traits).

- **The length of time for which a piece of data remains relevant**. The period over which data remains relevant, or when it may need to be 'refreshed', is driven by the frequency with which data points may change. For example, someone's browsing history may represent a useful data point for only a few minutes to several days, whereas their date of birth is relevant for their entire life.
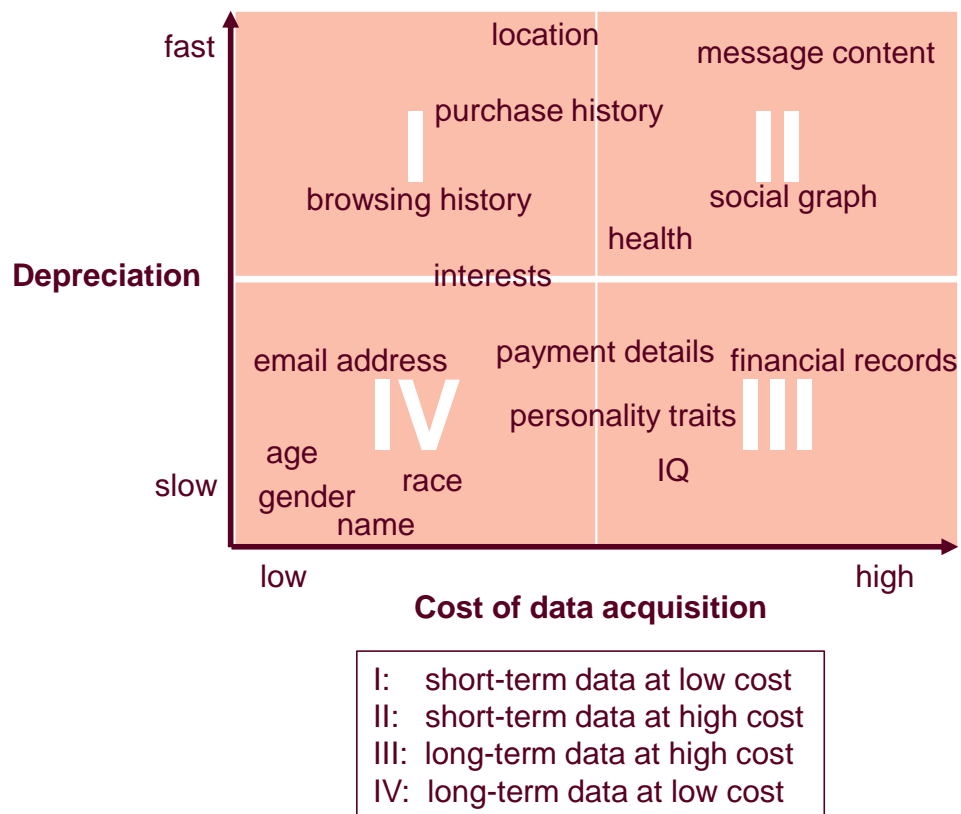
Figure 3.1 shows where some types of frequently used personal data could lie along these two dimensions. Starting in the bottom-left corner, data on demographics such as age tends to be widely available, as consumers can provide it multiple times and tend do so without much hesitation. Age also evolves in a fully predictable way; therefore, knowing a person's age once is sufficient for future reference. In contrast, browsing history is also being tracked by multiple firms at the same time, but it changes constantly and needs to be frequently updated to have any value.[8]

---

[6] Some data is not linked to individuals at all, such as weather information; however, a large set of data contains information on individuals in some way (e.g. when recording number plates).
[7] Data can be categorised in many ways; however, these two dimensions capture many of the aspects discussed elsewhere. For example, the discussion of whether data has properties of a public good (by being non-excludable and non-rivalrous) revolves around the question of whether datasets can be replicated. Ultimately, replicability is one factor affecting the cost of data collection. See, for example, Duch-Brown, N., Martens, B. and Mueller-Langer, F. (2017), 'The economics of ownership, access and trade in digital data', European Commission JRC Digital Economy Working Paper 2017-01, 17 February.
[8] Bujlow, T., Carela-Español, V., Solé-Pareta, J. and Barlet-Ros, P. (2017), 'Web Tracking: Mechanisms, implications, and Defenses', *Proceedings of the IEEE*, **105**:8, 28 July, pp. 1476–1510; Englehardt, S. and Narayanan, A. (2016), 'Online tracking: A 1-million-site measurement and analysis', *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, October, pp. 1388–1401.

**Figure 3.1    Characteristics of types of personal data**



Note: In this context, 'depreciation' refers to how fast the value of data erodes over time.

Source: Oxera.

Someone's social network and interactions (their 'social graph') is more likely to be accessible to only a few firms, as it is relatively costly to collect and requires regular updating. Complex inferred data, such as personality traits, may be available (in different forms) to various firms at different costs, as this information can be inferred from a range of factors. For example, online browsing behaviour and even bank transaction data can reveal certain personality traits, such as conscientiousness or extroversion.[9]

Firms with access to more comprehensive datasets are likely to have more accurate data—for example, computer models based only on Facebook 'likes' are reported to be more accurate at judging personality traits than friends and family.[10] The importance of the marginal impact of enhanced accuracy is likely to depend on the specific way it is used.

---

[9] Adeyemi, I. R., Abd Razak, S. and Salleh, M. (2016), 'Understanding Online Behavior: Exploring the Probability of Online Personality Trait Using Supervised Machine-Learning Approach', *Frontiers in ICT*, **3**:8, 31 May; Gin, J. (2017), 'Commercial Psychographic Personalisation', DataSine blog post, 15 November, https://datasine.com/2017/11/15/commercial-psychographic-personalisation, accessed 28 September.
[10] Youyou, W., Kosinski, M. and Stillwell, D. (2015), 'Computer-based personality judgments are more accurate than those made by humans', *PNAS*, **112**:4, 12 January, pp. 1036–40. See also the seminal paper Kosinski, M., Stillwell, D. and Graepel, T. (2013), 'Private traits and attributes are predictable from digital records of human behavior', *PNAS*, **110**:15, 12 February, pp. 5802–05.

# 4 Competition in data-driven markets

The use of data can affect competition. For example, concerns have been raised that a high concentration of data residing with a few firms could represent a barrier to entry.[11] This has drawn the attention of regulators and competition authorities, in particular whether certain practices prevent consumers from multi-homing and leverage data from one market to develop or improve products in other markets.

## 4.1 Privacy and competition

Privacy can be a dimension of competition, but it can also create tensions with competition aims, as making personal data widely accessible would directly hurt consumers. What is more, privacy preferences and consumers' definitions of privacy vary greatly across individuals and contexts,[12] which makes it challenging for uniform regulation to suit all consumers.EU policymakers have taken a first step to improve privacy outcomes by introducing the EU General Data Protection Regulation. Changes to the Privacy and Electronic Communications Regulations are on the horizon.[13]

## 4.2 Competition framework

The impact of data on competition is driven by the characteristics of the data itself and the importance of the data for the business model in question. Data that is cheap to obtain and that does not erode quickly in value is likely to be easily acquired by many firms, suggesting that they can compete effectively; meanwhile, data that does not erode quickly in value, but is costly to obtain, may enable a longer-term advantage in a specific data segment. These characteristics of specific pieces of data interact with the characteristics of the market, such as network effects and multi-homing.[14] For example, a lower cost of data acquisition is likely to be associated with more multi-homing.

Figure 4.1 describes the effect of the use of data and market characteristics on market outcomes for consumers.
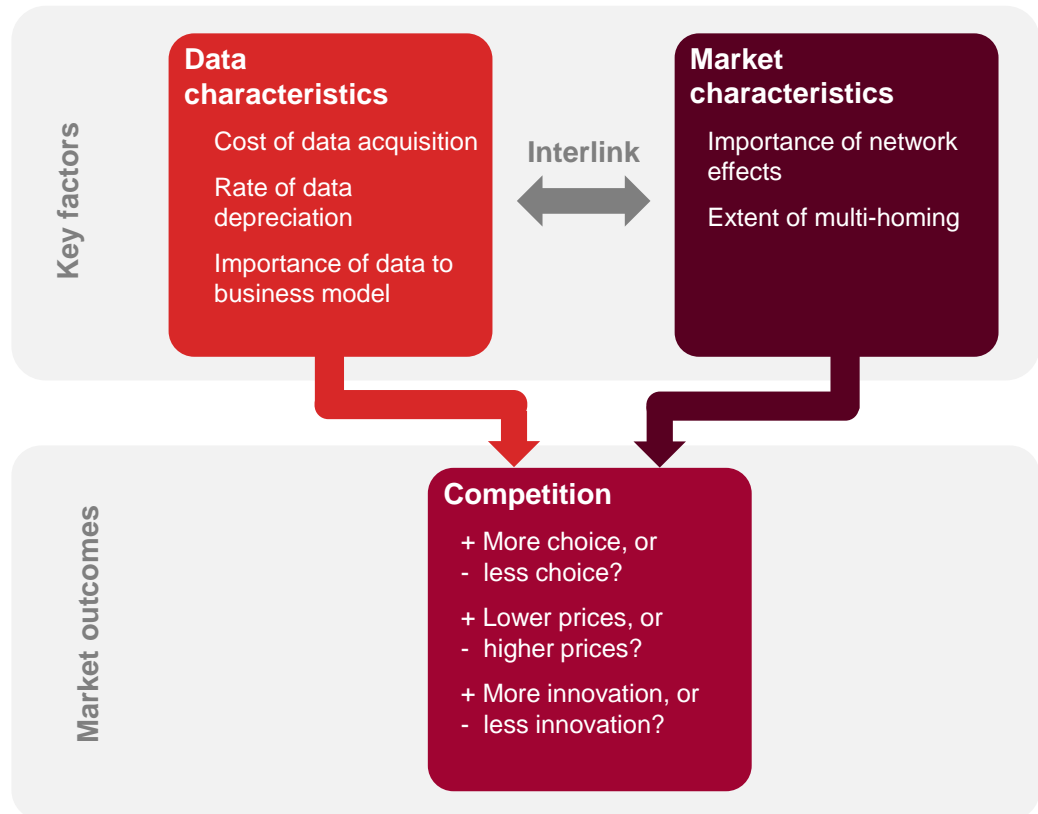
---

[11] See, for example, Kennedy, J. (2017), 'The Myth of Data Monopoly: Why Antitrust Concerns About Data Are Overblown', Information Technology and Innovation Foundation, 6 March, p. 7.

[12] For an overview, see Acquisti, A., Taylor, C., and Wagman, L. (2016), 'The Economics of Privacy', *Journal of Economic Literature*, **54**:2, section 3.8, pp. 442–492.

[13] Regulation (EEA) 2016/769 and The Privacy and Electronic Communications (EC Directive) Regulations 2003.

[14] A network effect is where the benefit that one consumer receives from a network product is affected by how many other consumers also use it. Network effects are a form of economies of scale driven by the demand characteristics of a product rather than the supply side. See Oxera (2013), 'Snowball effects: competition in markets with network externalities', *Agenda*, December. Multi-homing is where consumers use multiple platforms/websites/apps/providers for the same purpose. Using multiple messaging apps is an example of multi-homing.

**Figure 4.1    Framework for assessing the impact of data on market outcomes**



Source: Oxera.

Competition in data-driven markets will be more intense the more firms have access to a similar dataset that they use to offer a similar product. All else equal, the presence of multiple firms is expected to increase choice and innovation for consumers and drive down prices. Whether more than one firm is able to collect a similar dataset depends on the characteristics of the data and those of the market in which the data is used. Strong network effects favour the emergence of a few or one large player, as does single-homing by consumers. In such a case, smaller competitors may struggle to obtain a dataset similar to that of an incumbent.

Established players may be able to influence market characteristics, for example, by taking steps to increase the cost of multi-homing or to strengthen network effects by integrating a product into a wider ecosystem. Data can play a role in this if, for example, it is made incompatible with competing services or if it feeds into a range of different products. However, in practice, it is often challenging to assess whether such changes are anti-competitive or have efficiency benefits (such as an improved user experience).

### 4.3    Data trading

Another option for competition in data-driven markets to increase is through data trade between firms. Little trade is being observed at the moment, which may be due to a number of factors:

- data collection is still evolving rapidly, with a wide range of data formats in use;

- firms may struggle to find out who offers or would be interested in buying access to data, and if they know, agreeing on a price could be a challenge given the lack of comparable interactions;

- there may be some legal uncertainty with regard to when privacy regulation such as GDPR applies.

Some transactions occur in areas where these obstacles can be overcome more easily—for example, with machine learning data catalogues on data marketplaces.[15]

Whether a firm has an incentive to trade their data depends on the closeness of the market in which the potential trading partner is active.[16]

- If the other firm is active in the same market, a firm will have little incentive to provide access to it data as it would directly strengthen its competitors.[17]

- If the other firm is active in a related market (in the same value chain or an indirect substitute), the willingness to trade data depends on whether the other firm's valuation of the data exceeds the value of exclusivity for the data-owning firm (e.g. if it plans to expand into related markets).

- If the markets are not related, trade is likely to be beneficial for both firms as the data-owning firm can extract part of the buying firm's valuation of the data.

Trade with data on unrelated markets (e.g. to train algorithms) would be desirable from both the firm and the consumer perspective. If firms are active on related markets, a case-by-case assessment is necessary to examine whether mandated data access would markets more efficient or reduce the incentives to collect data in the first place. Such an assessment may be easier when data markets become more 'liquid' as some firms may decide to provide access without any intervention. In the case of direct competition, it is important to examine where competing firms cannot get direct access to data from consumers or other sources, whether the lack of access constitutes a high barrier to entry, and whether an obligation to share data would incentivise the firms to stop collecting data.

---

[15] See Smith, J. (2018), 'Data Marketplaces: The Holy Grail of our Information Age', Hacker Noon, 25 July, https://hackernoon.com/data-marketplaces-the-holy-grail-of-our-information-age-1211a6fec390, accessed 28 September.
[16] This assumes the data-owning firm has a monopoly over some of its data.
[17] This is the case unless the data-owning firm agrees a price with the other firm that exceeds its own valuation of the data.
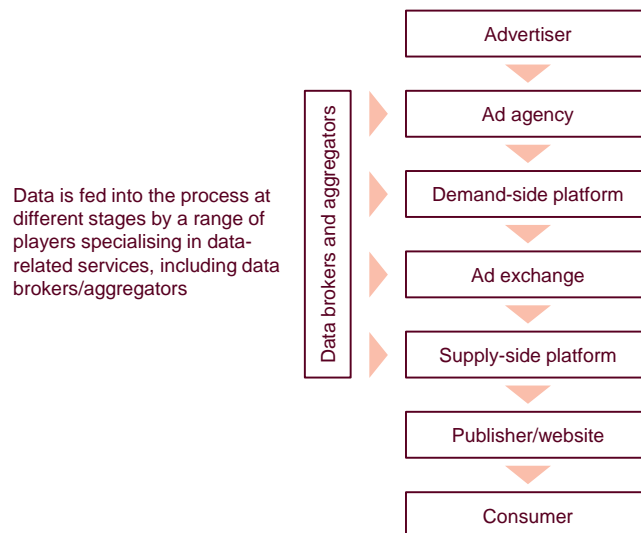
# 5    Case study on targeted advertising

One of the main uses of online data is in producing targeted advertising. Effective online advertising relies on consumer data to target relevant ads to individuals. This section describes the structure of online advertising markets and identifies potential areas of interest from the perspectives of both competition and privacy.

## 5.1    Industry structure

The digital advertising industry has multiple layers with many different players at each level (see Figure 5.1  Online advertising value chain). In general, advertisers—often through advertising agencies and demand-side platforms— bid for a suitable placement and target audience for their advertisement. Supply-side platforms aggregate the advertising opportunities and place them for auction on the advertising exchange. The matching of advertisements and audience ultimately happens on the advertising exchange.

**Figure 5.1    Online advertising value chain**



Source: Oxera.

There is some ambiguity with regard to the market definition of the final products, i.e. the ads delivered to consumers. The market definition is important in order to identify which firms exert competitive pressure on each other. For example, 'search' is often separated from 'display' advertising on the basis that a search term provides a clear indication of what a consumer is looking for.[18] In contrast, other studies highlight the possibility of digital and non-digital advertising converging, given new options of targeting in traditionally non-digital channels.[19] This would mean that a key advertising market would include both digital and TV marketing, potentially implying smaller importance of individual firms than if 'search' advertising was considered a market in itself.
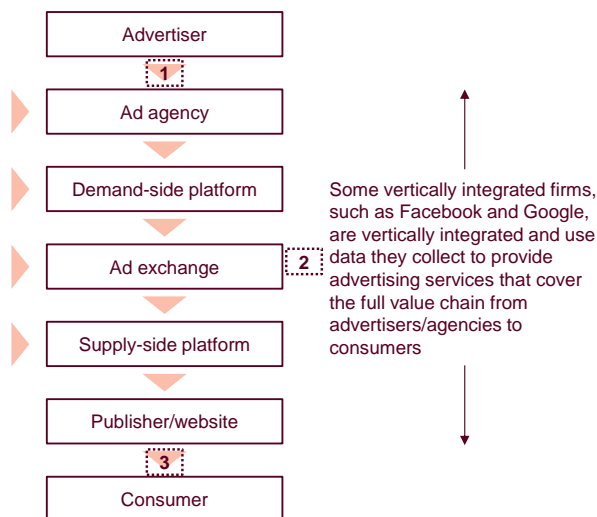
---

[18] For example, see Bundeskartellamt (2018), 'Online advertising: Series of papers on "Competition and Consumer Protection in the Digital Economy"', February.
[19] Monopolkommission, (2015), 'Competition policy: The challenge of digital markets', Special Report No. 68, 1 June, pp. 42–44.

## 5.2    Competitive issues

Competition in online advertising has come under scrutiny in various jurisdictions, including Australia,[20] France[21] and Germany.[22] Competition authorities have developed different theories on how some players could hinder competition, most of which relate to the strong and growing market position of Facebook and Google. Figure 5.2          Potential issues in online advertising illustrates where potential issues arise along the value chain.

**Figure 5.2    Potential issues in online advertising**



Source: Oxera.

Points 1, 2 and 3 highlight where a lack of competition could arise.

1.  A dominant position may allow a large, vertically integrated platform to set higher prices and/or offer lower quality to advertisers—for example, by creating scarcity on websites.[23]

2.  A dominant position may allow the platform to foreclose competitors from access to their systems—for example, by limiting interoperability with upstream or downstream 'partial' competitors.

3.  A dominant position may allow a platform to impose privacy terms to consumers that would not be acceptable under greater competition;[24] however, some market observers consider this a possibility for a wider group of suppliers also in the absence of market dominance.

Consumer data is used to make adverts relevant. Vertically integrated advertising platforms often rely on large amounts of user data from their websites and apps, and they may not make this data available to advertisers. This has raised a debate regarding the replicability of these datasets, as other parties may have to incur high costs to obtain the same data.

---

[20] See Australian Competition & Consumer Commission (2018), 'Digital Platforms Inquiry', 26 February, https://www.accc.gov.au/about-us/inquiries/digital-platforms-inquiry, accessed 28 September.
[21] Autorité de la Concurrence (2018), 'Avis n° 18-A-03 du 6 mars 2018 portant sur l'exploitation des données dans le secteur de la publicité sur internet', 6 March.
[22] Bundeskartellamt (2018), 'Bundeskartellamt launches sector inquiry into market conditions in online advertising sector', Press release, 1 February.
[23] Monopolkommission, (2015), 'Competition policy: The challenge of digital markets', Special Report No. 68, 1 June, pp. 46–7.
[24] Bundeskartellamt (2016), 'Bundeskartellamt initiates proceeding against Facebook on suspicion of having abused its market power by infringing data protection rules', press release, 2 March.

# 6    Conclusions

Companies have access to far more data than they have ever had before. Such access has raised concerns about market outcomes, including about privacy. However, it has also led to positive changes in many markets and sectors across the economy. Many business models rely on data to facilitate transactions, to generate revenues through targeted advertising, and to automate complex processes that feed into a range of new products (such as IoT devices). Through these, data has provided consumers with new products and services and made existing products and services better and cheaper.

The economic characteristics of data—particularly its cost of acquisition and rate of depreciation—determine whether multiple firms can get access to a similar dataset. The replicability of data is important to assess whether data may create competitive concerns.

Data has become a competitive variable in many digital markets, drawing the attention of competition authorities. Privacy and competition interact in relation to personal data hence a cohesive approach is required when considering any interventions. Both the characteristics of data and of the market in which it is used determine whether concerns are likely to arise. Data trading, while still limited, may allow firms to exploit data more efficiently in the future.