

Access to big data as a remedy in big tech

Nick Dadson, Iain Snoddy and Joshua White*

Abstract: ‘Big data’ and ‘big tech’ have become central topics in recent antitrust debate and regulation. For example, the Competition and Markets Authority (CMA) recently published a report on online platforms, expressing concerns that the major platforms like Google are now protected from competition by such strong incumbency advantages. Underlying the CMA’s theory of harm is the essential facility theory of antitrust, under which Google’s ability to control access to its click-and-query data is seen as preventing its rivals from competing effectively. EU jurisprudence has identified three criteria to determine whether data are an essential facility and whether access should be mandated. First, the data must be indispensable to compete in the market. Secondly, absent data sharing, technical improvements by competitors must be hampered or precluded. Thirdly, there must be no objective justification to refuse competitors access to the data. It is difficult to reconcile the authorities’ concerns with Google’s click-and-query data with these criteria, however. Actual and potential alternatives exist; Google’s competitors have been innovating in the search market for more than a decade; and there are objective reasons to limit data access, including threats to innovation and privacy concerns.

Keywords: Competition and Markets Authority, CMA, data access, essential facility theory, online platforms, digital advertising, innovation, privacy

1. Introduction

Recent technological advances have lowered the costs of collecting, storing and analysing large amounts of data. As a consequence, the amount of and use of data in the economy has exploded. Companies with internet-connected products can often directly collect information on their customers, such as their usage patterns and GPS locations.¹ Machine learning algorithms have helped companies use data to tailor their products to individual customers, evaluate sales performance or monetize their user base by selling advertising. Other companies have directly monetized their data by selling them to other companies.

As data have increased in relevance, ‘big data’ and ‘big tech’ have become central topics in recent antitrust debates, and tech companies have become a focus of

regulators’ and politicians’ attention.² For example, the Competition and Markets Authority (CMA) recently published a report on online platforms and digital advertising in the UK (the ‘CMA Report’).^{3, 4} The CMA Report focused on Google and Facebook, and expressed concerns that the two platforms ‘are now protected by such strong incumbency advantages – including network effects, economies of scale and unmatched access to user data – that potential rivals can no longer compete on equal terms’.⁵

The CMA Report both identified potential antitrust concerns and proposed various remedies to ameliorate the concerns it identified. In this article, we outline their proposed Google data access remedy whereby a new or set of already existing institutional bodies would be empowered to implement various regulatory functions.

* Nick Dadson is a Manager and Iain Snoddy is an Associate at Analysis Group in Montreal, Canada. Joshua White is a Vice President at Analysis Group in London, UK. The views presented in this article are those of the authors and do not represent any opinions or positions of Analysis Group or its affiliated companies. The authors did not receive financial support for their work on this article.

1 A. Hagiu and J. Wright, ‘When Data Creates Competitive Advantage’, *Harvard Business Review* (January–February 2020).

2 C. Kang and D. McCabe, ‘Lawmakers, United in Their Ire, Lash Out at Big Tech’s Leaders’, *The New York Times* (1 July 2020), available at: <https://www.nytimes.com/2020/07/29/technology/big-tech-hearing-apple-amazon-facebook-google.html> (accessed 10 February 2021).

3 Competition and Markets Authority, ‘Online platforms and digital advertising market study’, gov.uk (1 July 2020), available at: <https://www.gov.uk/cma-cases/online-platforms-and-digital-advertising-market-study> (accessed 10 February 2021).

4 Similarly, the European Commission’s recently proposed Digital Markets Act (DMA) proposes criteria for classifying online platforms as ‘gatekeepers’, as well as the obligations that designated gatekeepers should comply with. See European Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)’, EUR-Lex (15 December 2020), available at: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN> (accessed 10 February 2021).

5 CMA Report, p 5.

Such a 'Digital Markets Unit' (DMU) could, for example, require Google to give rival third-party search engines access to Google Search Click and Query (or 'click-and-query') data.⁶ Click-and-query data contain information on what users searched for and how they interacted with the results provided by the search engine.⁷ According to the CMA, their proposed remedy is intended to provide access to data which would arguably improve the quality of rivals' search algorithms and increase competition in digital advertising.⁸ Underlying the CMA's theory of harm and the proposed remedy is the essential facility theory of antitrust in which Google's ability to control access to its click-and-query data prevents its rivals from competing effectively. The proposed remedy to ensuing consumer harm is to provide access to the facility deemed essential to competition.⁹

Academic and regulatory communities still debate how data may confer market power and the appropriate role of antitrust enforcement and regulation in that sphere. Evaluating data-driven market power and the regulation of data access is difficult. Many data-intensive industries are new and operating in a rapidly evolving competitive landscape where data are plentiful and already significantly fragmented. Data can be central to innovation, and privacy issues introduce significant tensions which impact any desire to share data more widely.

EU jurisprudence has identified three criteria to determine whether data are an essential facility and whether access should be mandated. First, the data must be indispensable to compete in the market. Secondly, absent data sharing, technical improvements by competitors must be hampered or precluded. Thirdly, there must be no objective justification to refuse competitors access to the data.¹⁰

It is difficult to reconcile the authorities' concerns with Google's click-and-query data with these criteria because actual and potential alternatives exist. Google's competitors have been innovating in the search market for more

than a decade and there are objective reasons to limit data access, including threats to innovation and privacy concerns.

In Section 2 we provide an overview of the CMA Report access remedy proposed for Google's click-and-query data. In Section 3 we discuss data access within the context of the essential facilities doctrine. In Section 4 we consider whether the first two essential facilities criteria are met for Google Search click-and-query data. In Section 5 we discuss the potential impact of granting competitors access to these data and in Section 6 we briefly conclude.

2. The CMA market study on online platforms

The CMA began a market study on online platforms and the digital advertising market in the UK in July 2019. The launch of this market study was triggered by concerns that online platforms in the digital advertising space, particularly Google and Facebook, had market power and that the 'sector [was] not working well'.¹¹ While the CMA never defined what 'not working well' might mean in terms of lost consumer welfare, it published an interim report in December 2019 and published responses to the interim report in April 2020.¹² The CMA Report, published on 1 July 2020,¹³ contained a number of concerns about Google's position in the Search Advertising Market¹⁴ as well as proposed remedies, including a proposed data access remedy to rectify its concerns.

2.1. The CMA Report's concerns with Google

As discussed in the CMA Report, Google offers search advertising in which advertisements can be targeted based on user click-and-query behaviour and personal

6 The CMA Report refers to a 'Digital Markets Unit'. See CMA Report, para 76. We do not comment on DMUs in this article.

7 CMA Report, paras 25, 3.65.

8 CMA Report, paras 48, 90, 125.

9 K. Hylton, 'Digital Platforms and Antitrust Law' (2019) 19-8 Boston University School of Law, Law and Economics Research Paper, available at: https://scholarship.law.bu.edu/faculty_scholarship/605/ (accessed 10 February 2021).

10 E. Bruc, 'Data as an Essential Facility in European Law: How to Define the "Target" Market and Divert the Data Pipeline?' (2019) 15(2-3) European Competition Journal 9.

11 Competition and Markets Authority (fn 3), para 6.

12 Competition and Markets Authority (fn 3).

13 Ibid.

14 There are three types of digital advertising according to the CMA: search, display and classified. The CMA defines search advertising as the sponsored ads provided in response to users' queries on a search engine. Advertisers pay for advertisements, e.g., a text link, to be displayed

when users enter specific keywords or phrases. Display advertising, according to the CMA, enables advertisers to place ads on websites or apps in a variety of formats. Classified advertising, unlike the other two types of advertising, involves advertisers paying online companies to list specific products or services on a specialized website. According to the CMA, search and display advertising perform different roles in the decision-making process of customers and therefore do not substitute for each other easily. The CMA also argues that there is limited substitutability between digital and traditional advertising media given that unlike traditional media, digital advertising allows advertisers to target specific audiences with data. We note that the CMA did not carry out a formal market definition exercise, but rather 'looked at competitive constraints across the sector'. See CMA Report, para 3.25. Nevertheless, despite not proposing a formal market definition, the CMA implicitly defines separate relevant markets for search advertising (the 'Search Advertising Market') and display advertising (the 'Display Advertising Market') and uses these in its assessment of potential harm and evaluation of remedies. See CMA Report, paras 2, 2.43, 2.44, 2.47, 2.53, 5.23.

user information.¹⁵ The CMA Report focuses on the static share of search advertising, noting in 2019 that Google accounted for more than 90% of search advertising revenues in the UK.¹⁶ The CMA Report goes on to suggest that six characteristics of Google and the digital advertising market may inhibit rivals' entry or expansion: (1) network effects and economies of scale; (2) consumer behaviour and the power of defaults; (3) unequal access to user data; (4) a lack of transparency;¹⁷ (5) the importance of ecosystems; and (6) vertical integration and conflicts of interest.¹⁸ We focus on the first and third characteristics in this article, which relate to Google's scale in, and access to, user data generated through users' searches via the Google search engine.

The CMA Report asserts that greater scale in click-and-query data is advantageous because it enables search engines to improve their algorithms, returning more accurate results to users and allowing better targeting for advertisers. It also argues that this advantage is higher for uncommon or new queries, i.e., 'tail queries'.¹⁹ Since tail queries make up a sizable proportion of the user queries seen, the CMA contends that the ability to produce more accurate or useful search results in response to tail queries is likely valuable to search engine users, driving increased engagement with the search engine.²⁰ Furthermore, the CMA Report suggests that advertisers also benefit from the ability to target their advertising with more precision and achieve higher returns on their investment.²¹ Finally, the CMA Report argues that Google collects a large amount of user data from its user-facing services, its Android mobile devices, and the tags they place on third-party websites and apps, and the ability to track user behaviour online enables Google to offer a number of valuable targeted-advertising options and measure the effectiveness of that advertising.²²

The CMA Report further notes that they analysed over three billion search events on Google and Bing during one week, and estimated that Google received 16 tail queries for every one tail query received by Bing.²³ In addition, according to the CMA, Google tags are found on over 80% of the most popular websites, a proportion much larger than any other platforms.²⁴ The CMA Report contends that Google's ability to acquire vast amounts of user data and (potential) rivals' lack of access to comparable data limits competition in the Search Advertising Market and leads to consumer harm.²⁵

2.2. The proposed data access remedy

In light of Google's position, the CMA Report recommends a two-part regulatory regime for online platforms that earn their revenues via digital advertising. The part first consists of high-level principles for platforms with 'Strategic Market Status',²⁶ intended to address the potential effects of market power rather the sources of it. In particular, the CMA recommends establishing an enforceable code of conduct based on 'fair trading', 'open choices' and 'trust and transparency' for platforms with 'Strategic Market Status'. The CMA does not provide any specific criteria to assess whether a platform should have SMS but mentions that a Digital Markets Taskforce 'will be advising government on the specific criteria for deciding whether a digital platform has SMS'.²⁷

The second part of the proposed remedy consists of measures to address the sources of market power in the advertising markets the CMA has implicitly defined. The CMA recommends specific interventions 'to overcome barriers to entry and expansion on the demand and supply side of search'.²⁸ They recommend creating the DMU and investing it with power to restrict Google's

15 'Google Safety Center', Google, available at: <https://safety.google/privacy/ads-and-data/> (accessed 10 February 2021).

16 CMA Report, para 18.

17 The CMA claims that the use of complicated algorithms by the online platforms has made it particularly challenging for market participants to understand and contest decisions taken by Google and the other platforms. The CMA also reports a substantial lack of transparency faced by market participants concerning multiple aspects of digital advertising, such as the quality and effectiveness of advertising and the functioning of auctions.

18 CMA Report, paras 23–30 (network effects and economies of scale); 31–40 (consumer behaviour and the power of defaults); 41–48 (unequal access to user data); 49–55 (lack of transparency); 56–60 (importance of ecosystems); and 61–68 (vertical integration and conflicts of interest).

19 'Tail queries' are uncommon, rare or new queries. The CMA Report defines tail queries formally (in their analysis of search events seen by Google and Bing in a one-week period in the UK) as the 30% of queries seen the least often. See CMA Report, paras 3.27, 3.68, and Figure 3.4.

20 CMA Report, para 3.76.

21 CMA Report, para 25.

22 CMA Report, paras 42–45.

23 CMA Report, para 3.69, and CMA Report Appendix I, para 46.

24 CMA Report, para 45.

25 CMA Report, paras 27, 48, 3.78.

26 CMA Report, paras 77–82. The CMA Report roughly defines 'Strategic Market Status' to designate firms 'that have obtained gatekeeper positions and have enduring market power over the users of their products'. See CMA Report, para 7.56.

27 CMA Report, para 7.58. The CMA Digital Markets Taskforce has recently set out its advice on the design and implementation of pro-competitive measures. These include introducing an enforceable code of conduct for the industry, engaging in pro-competitive interventions, and increasing the scrutiny of firms designated with strategic market status during merger investigations. See Competition and Markets Authority, 'A New Pro-Competition Regime for Digital Markets: Advice of the Digital Markets Taskforce', gov.uk (8 December 2020), available at: <https://www.gov.uk/cma-cases/digital-markets-taskforce#taskforce-advice> (accessed 10 February 2021).

28 CMA Report, para 88.

ability to secure default positions across multiple devices and browsers,^{29, 30} and to 'require Google to provide click-and-query data to third-party search engines to allow them to improve their search algorithms'.^{31, 32} These data would include information on some or all user queries, user interactions – such as clicks, locations and previous searches – and search results.³³ The focus of the remainder of the article is on this digital access remedy, and the CMA's claims that it would reduce Google's scale advantages over actual and potential rivals with respect to data.³⁴

3. Data as an essential facility

With its focus on potential barriers to entry stemming from the inability of rivals to obtain click-and-query data comparable to those received by Google, the CMA is implicitly applying the essential facilities doctrine in its assessment of the competitive conditions in the digital advertising market and its proposed remedies. While the application of the essential facilities doctrine can vary substantially across countries,³⁵ in general a facility or infrastructure is considered 'essential' if: (1) it is used extensively in business operations, and (2) its duplication is considered difficult, if not impossible, due to legal, economic, physical or geographic constraints.³⁶ If a facility is truly essential, then a firm controlling the facility will be immune to most forms of competition.³⁷ In these cases, the essential facilities doctrine mandates

that to ensure competition, the controlling firm must provide access to the facility at a reasonable price.

The use of the essential facilities doctrine dates back to the US Supreme Court's *Terminal Railroad* decision in 1912.³⁸ In this decision, the Court mandated that the association of railroads that controlled all railroad access through St. Louis provide competing railroad carriers access to its facilities at reasonable terms.³⁹ Since *Terminal Railroad*, the essential facilities doctrine has been applied to seminal cases in the US,⁴⁰ but more recently its use has declined due to an increasing aversion to regulating the prices of and granting access to 'facilities' created through competition among players.⁴¹ In particular, the *Trinko* decision in 2004 significantly narrowed the applicability of the essential facilities doctrine⁴² by restricting its use to situations where access was strictly unavailable. Furthermore, it required the plaintiff to show that there is no legitimate or procompetitive rationale to deny access; the evidence needs to convincingly show that the dominant firm seeks to eliminate a rival.⁴³ Since *Trinko*, the essential facilities doctrine has rarely been successfully invoked in the US.⁴⁴

In comparison with the US, over the past 30 years EU judges and authorities have been more receptive to the essential facilities doctrine.⁴⁵ Essential facility issues fall under Article 102 of the Treaty on the Functioning of the European Union (TFEU), which covers situations where a dominant firm refuses to supply its products or services; refuses to grant access to its facilities; or

29 The CMA uses the term 'Digital Markets Unit' to refer to an institutional body empowered to implement the regulatory functions they recommend. This could be a new or an existing (set of) institution(s). See CMA Report, para 76.

30 CMA Report, para 89.

31 CMA Report, para 90.

32 The EC's proposed DMA also contains several provisions related to data portability and access, but these proposals are more limited and generally are restricted to providing end users and businesses with access to their own data. The DMA does include a provision for requiring rival search engines to be allowed access to click-and-query data on fair, reasonable and non-discriminatory terms; however, the data will be subject to anonymisation of query, click and view information containing personal information. DMA, Articles 6(i)–(j).

33 CMA Report Appendix V, paras 92–97.

34 CMA Report, paras 90, 8.43.

35 'The Essential Facilities Concept' (OECD, 1996), available at: <http://www.oecd.org/competition/abuse/1920021.pdf> (accessed 10 February 2021).

36 Directorate-General for Competition, *Glossary of terms used in EU competition policy: Antitrust and control of concentrations* (European Commission, 2002), p 18.

37 D. Patterson, R. Pitofsky and J. Hooks, 'The Essential Facilities Doctrine under United States Antitrust Law' (2002) 70 *Antitrust Law Journal* 443.

38 *United States v. Terminal Railroad Association of St Louis*, 224 US 383, 410 (1912).

39 The Supreme Court noted that it was almost impossible for competing carriers to develop their own facilities, and that it was necessary for carriers to pass through the St. Louis railroads to access railways surrounding

the city. For details, see M.A. Piropatto, 'Open Access and the Essential Facilities Doctrine: Promoting Competition and Innovation' (2000) 1 *University of Chicago Legal Forum* 380–81; Patterson, Pitofsky and Hooks (fn 37), pp 445–46.

40 In *MCI Communications*, for example, the primary question was whether AT&T needed to open its network control and its effective monopoly in long-distance and local telephone markets to competition. The Seventh Circuit Court in *MCI Communications* stipulated specific requirements necessary to establish an essential facilities claim. A firm with monopoly power violates section 2 of the Sherman Act when the monopolist controls access to an essential facility, it is not possible to duplicate the facility, the monopolist denies competitors access to the facility, and the monopolist is reasonably able to provide access to the facility. While the court ruled that AT&T's provision of interconnection for multipoint service was not an essential facility, it ruled in favor of giving MCI access to AT&T's local service network. See *MCI Communications Corporation v. AT&T Co.*, 708 F.2d 10801, 1132–33 (7th Cir. 1983).

41 Bruc (fn 10), p 7.

42 *Verizon Communications Inc. v. Law Offices of Curtis V. Trinko*, 124 S. Ct. 872 (2004), rev'g 305 F.3d 89 (2d Cir. 2002) Part III.; F.X. Schoen, 'Exclusionary Conduct after *Trinko*' (2005) 80(5) *New York University Law Review* 1625.

43 Hylton (fn 9).

44 S. Weber Waller and W. Tasch, 'Harmonizing Essential Facilities' (2010) 76(3) *Antitrust Law Journal* 745; Bruc (fn 10), pp 6–7.

45 G. Colangelo and M. Maggolino, 'Big Data as Misleading Facilities' (2017) 13(2–3) *European Competition Journal* 249–81.

sets prices, delivery and quality in a way that has the same effect as a refusal.⁴⁶

The essential facilities doctrine was applied broadly throughout the EU following the break-up and privatization of state-owned monopolies in the 1990s.⁴⁷ The European Commission (the 'EC' or 'Commission') first referred to the doctrine in *Sealink*, in which the European Commission Directorate General for Competition (DG Comp) determined that the Holyhead port where Stena Sealink owned port facilities and refused access was an essential facility, since it was the only British port serving the market.⁴⁸ The doctrine has since been used in multiple court and National Competition Authority decisions in which the owners of ports, harbours, tunnels and similar facilities were accused of using their control of an essential input to prevent downstream competition.^{49, 50}

The doctrine has been applied to cases involving intellectual property as well. In *Magill*, the European Court of Justice (CJEU) confirmed an earlier Commission decision stating that withholding access to television listing information could constitute an abuse of a dominant position. In its judgment, the CJEU set out specific requirements for the antitrust analysis of the refusal to supply intellectual property. In applying these criteria, the Court determined that: (1) the television listings were indispensable; (2) there was no justification for refusing to supply the listing; (3) refusal to supply prevented the appearance of a new product; and (4) refusal to supply effectively gave broadcasters a de facto monopoly in a secondary, downstream market of weekly television guides.⁵¹

Subsequent decisions, namely *Bronner*⁵² and *IMS Health*,⁵³ confirmed the use of specific criteria to determine whether access should be granted.^{54, 55} In *IMS Health*, in particular, the CJEU found that each of the 'exceptional circumstances' stipulated in *Magill* need to be satisfied for access be granted.⁵⁶ In *Microsoft*,⁵⁷ the CJEU subsequently broadened the third test (does the refusal to supply prevent the appearance of a new product?) to include cases where the refusal may prevent the development of improved products with added value.⁵⁸

In each of these decisions, the legal decisions were consistent with a coherent set of underlying economic concerns, namely that the competitors could not generate incremental consumer welfare by developing their own alternative to the essential facility, and that this resulted in monopoly rents to the owner of the essential facility which further decreased social welfare. Further, in none of these cases was any objective or pro-competitive justification identified which would mitigate the reduction in consumer welfare arising from refusing access.

4. Do Google click-and-query data constitute an essential facility?

As previously mentioned, under the essential facilities doctrine, the exercise of intellectual property rights alone is not an abuse of dominance, unless: (1) the data are indispensable to compete in the market,⁵⁹

46 Bruc (fn 10), pp 7–8.

47 S. Baches Opi, 'The Application of the Essential Facilities Doctrine to Intellectual Property Licensing in the European Union and the United States: Are Intellectual Property Rights Still Sacrosanct?' (2001) 11(2) *Fordham Intellectual Property, Media & Entertainment Law Journal* 418.

48 *Ibid*, pp 429–430.

49 Weber Waller and Tasch (fn 44), p 741.

50 See e.g. *Attheraces Ltd. v. British Horseracing Bd. Ltd.* [2005] EWHC (Ch) 3015 (Eng.), *Software Cellular Network Ltd. v. T-Mobile (UK) Ltd.* [2007] EWHC (Ch) 1790, 58 (Eng.) and Case COMP/C-3/37.792 *Microsoft*.

51 Bruc (fn 10), p 8; Colangelo and Maggolino (fn 45), pp 249–81; J. Turney, 'Defining the Limits of the EU Essential Facilities Doctrine on Intellectual Property Rights: The Primacy of Securing Optimal Innovation' (2005) 3(2) *Northwestern Journal of Technology and Intellectual Property* 187.

52 *Oscar Bronner GmbH & Co. KG v. Mediaprint Zeitungs- und Zeitschriftenverlag GmbH & Co. KG*, Case C-7/97, 1998 E.C.R. I-7791, [1999] 4 C.M.L.R. 112.

53 Case C-481/01, *IMS Health GmbH & Co. OHG v. NDC Health GmbH & Co. KG*, 2004 O.J. (C3) 16 (29 April 2004).

54 In *Bronner*, for example, '[t]he Court decided these conditions were not met since other, less advantageous methods of distributing daily newspapers existed, and there were no technical, legal, or economic obstacles to establishing another nationwide delivery scheme... The Court held that it was not sufficient that the establishment of a second home-delivery scheme was not economically viable because of the small circulation of the daily newspaper to be distributed. Instead, it said that it must be demonstrated that it is not economically feasible to create a second delivery scheme for the distribution of daily newspapers with a circulation comparable to that

of the daily newspapers distributed by Mediaprint'. See S.J. Evrard, 'Essential Facilities in the European Union: Bronner and Beyond' (2004) 10 *Columbia Journal of European Law* 13.

55 *IMS Health* generates reports based on sales and prescription data purchased from pharmaceutical wholesalers. The data are checked and formatted according to a geographic segmentation of Germany into 1,860 areas, the '1860 brick structure'. Pharmaceutical companies would use this data to focus their marketing efforts on doctors located in certain areas. DG Comp determined that the brick structure was indispensable as it had become standard in the pharmaceutical industry, and that *IMS Health* was (unjustifiably) using its right to the brick structure to eliminate competition. It ordered that *IMS Health* grant licenses for their competitors to use the brick structure. *Ibid*, pp 25–57.

56 The CJEU emphasized the requirement that a third-party bring a new product to the market; the Court considered this requirement to be met even if the market was just potential or hypothetical. Therefore, it is no longer necessary to prove that the requested input has been traded as an independent product by the holder. See Colangelo and Maggolino (fn 45); Bruc (fn 10), p 12.

57 General Court EU, 17 September 2007, Case T-201/04.

58 Colangelo and Maggolino (fn 45).

59 The criterion is often broken up into two, closely related criteria: (1) the data are indispensable; and (2) refusing access to the data will eliminate all effective competition. These criteria are so closely related that some scholars have suggested they are really one criterion. For ease of exposition, this is the approach we follow here. See e.g. Evrard (fn 54), p 18: 'It seems that if a facility is essential for the requesting undertaking, the refusal to grant access to it will inevitably prevent that undertaking from

(2) the refusal restrains technical improvement, and (3) there is no objective justification for refusal.⁶⁰

4.1. Are Google Search click-and-query data indispensable to advertisers who compete in the digital advertising market?

Google Search click-and-query data are indispensable only if there are no alternatives available to advertisers. However, alternatives are easily identifiable.

First, a number of other search engines currently operate in the Search Advertising Market.⁶¹ While the scale of the data collected by Google Search maybe larger, the CMA has provided no evidence that the smaller search engines with access to millions or billions of data points do not provide adequate substitutes, particularly given the availability and continuing development of sophisticated algorithms, sampling strategies and extrapolation allowing them to make better use of this data.⁶²

Secondly, Google Search competitors currently collect a substantial amount of user data, and it is possible for them to collect more.⁶³ Companies can acquire additional click-and-query data from other players active in the supply of search (or digital) ad inventory. For example, in 2005 Yahoo acquired search and advertising company Whereonearth in an effort to improve search for both users and advertisers.⁶⁴ Furthermore, in 2010 Microsoft effectively purchased Yahoo's search and

search advertising businesses, along with the click-and-query data that Yahoo had collected.⁶⁵ More generally, a number of big datasets can be purchased from data vendors that can be integrated with existing data controlled by the competing search engines.⁶⁶ Moreover, advances in data science allow for competing search engines to generate better quality data themselves. For example, Microsoft Bing recently developed a novel artificial intelligence (AI) approach for generating large amounts of high-quality labelled data with which to train their machine learning algorithms.⁶⁷

Thirdly, indispensability requires a forward-looking assessment.⁶⁸ In addition to the current alternatives to Google Search, potential alternatives may be developed. Innovative companies often find novel ways to use existing data differently or to obtain data from different sources to compete. For example, while a manufacturer of smart refrigerators may use its data to predict individuals' dietary habits, a supermarket may try to do the same through an analysis of credit card traces.⁶⁹ Similarly, while Google Search data may provide insights into consumer preferences, there are a number of social networks that can provide similar information. Google Search itself demonstrated the ability to, paraphrasing the CJEU in *IMS Health*, viably produce a similar facility.⁷⁰ Despite initially lacking the scale of its competitors, Google used their PageRank algorithm to improve the speed of Isearch and surpass the incumbent search engine Yahoo – just as Yahoo had surpassed competitors

competing on the market and, thus, will eliminate it. Alternatively, if the refusal to use the facility is not likely to eliminate all competition on the part of the requesting undertaking, it inevitably means that the facility is not essential.'

60 See Evrard (fn 54), p 12; and Bruc (fn 10), p 9.

61 Indeed, outside of Google, Yahoo and Bing, there are hundreds of search engines, and consumers' ability to easily switch among search engines constrains Google's ability to behave anticompetitively. See R. Telang, U. Rajan and T. Mukhopadhyay, 'The Market Structure for Internet Search Engines' (2004) 21(2) *Journal of Management Information Systems* 138; R.H. Bork and J.G. Sidak, 'What Does the Chicago School Teach About Internet Search and the Antitrust Treatment of Google?' (2012) 8(4) *Journal of Competition Law & Economics* 671; M.A. Salinger and R.J. Levinson, 'Economics and the FTC's Google Investigation' (2015) 46 *Review of Industrial Organization* 33; 'How Search algorithms work', Google, available at: <https://www.google.com/search/howsearchworks/algorithms/> (accessed 10 February 2021).

62 See Y. Zhan and C. Ling, 'A strategy to apply machine learning to small datasets in materials science' (2018) 4(1) *npj Computational Materials* 1–8; H.J. Wilson and P.R. Daugherty, 'Small Data Can Play a Big Role in AI', *Harvard Business Review* (17 February 2020), available at: <https://hbr.org/2020/02/small-data-can-play-a-big-role-in-ai> (accessed 10 February 2021).

63 Bork and Sidak (fn 61), p 682 ('Microsoft, Yahoo, and Ask.com all offer search engines with significant user volume. In March 2012, there were more than 500 million searches on Ask.com, more than 2.5 billion searches on Yahoo, and more than 2.8 billion searches on Microsoft sites. Each of these competitors can direct consumers to a website in the same manner as Google').

64 'Yahoo! and Whereonearth Get More Relevant Together', Yahoo Search Blog (18 October 2005), available at: <https://web.archive.org/web/20080423224942/http://www.ysearchblog.com/archives/000199.html> (accessed 10 February 2021).

65 See e.g. J. Newman, 'Will Bing's Back-End Make Yahoo Better?', *PCWorld* (25 August 2010), available at: https://www.pcworld.com/article/204134/Will_Bing_Back_End_Make_Yahoo_Better.html (accessed 10 February 2021); J. Newman, 'Yahoo-Microsoft Deal Makes Bing Better, DOJ Says' (10 February 2010), available at: https://www.pcworld.com/article/189801/yahoo_microsoft_deal_makes_bing_better_doj_says.html (accessed 10 February 2021); J.E. Vascellaro and N. Wingfield, 'Microsoft, Yahoo Reach Search Deal', *The Wall Street Journal* (30 July 2009), available at: <https://www.wsj.com/articles/SB124886852386589989> (accessed 10 February 2021).

66 A. Lambrecht and C.E. Tucker, 'Can Big Data Protect a Firm From Competition?' (2017) *CPI Antitrust Chronicle* 2.

67 K.-H. Lee, X. He, L. Zhang and L. Yang, 'CleanNet: Transfer Learning for Scalable Image Classifier Training with Label Noise', *Arxiv.org* (25 March 2018), available at: <https://arxiv.org/pdf/1711.07131.pdf> (accessed 10 February 2021); 'Artificial Intelligence + Human Intelligence: Training Data Breakthrough', *Bing Blogs* (18 June 2018), available at: <https://blogs.bing.com/search-quality-insights/2018-06/Artificial-intelligence-human-intelligence-Training-data-breakthrough> (accessed 10 February 2021).

68 Evrard (fn 54), p 23.

69 Colangelo and Maggolino (fn 45).

70 In *IMS Health*, the European Court of Justice 'repeated the assessment contained in the [Bronner] judgment that it should not be economically viable for a firm of comparable size to the right holder to produce a similar facility'. See Turney (fn 51), p 193.

before it.⁷¹ In the words of Colangelo and Maggolino, '[i]n short, exclusivity over data does not entail exclusivity over the specific pieces of information needed'.⁷²

Fourthly, Google Search has significant competition outside the Search Advertising Market as implicitly defined in the CMA Report.⁷³ Advertisers have a lot of options for engaging with consumers outside of direct search advertising, and therefore the existence of these other substitutes serves as a competitive constraint on players in the Search Advertising Market. As such, market definition, which is already very difficult for search engines and other multi-sided platforms, is even more difficult in the digital advertising space.⁷⁴ Google faces competitive constraints from a number of sources, including, but not limited to, vertical websites, specialist shopping sites, online display advertising⁷⁵ and offline advertising.

'Vertical' websites, for example, allow users to perform specialized searches for specific types of information and offer an alternative platform to promote brands and products online. The CMA acknowledges the existence of these websites, but argues that general search engines can be used to reach different audiences:

General search has a higher reach, is cheaper and is best used to attract traffic from a wider audience that has demonstrated general intent, while specialized search is more expensive, has lower reach and is used to directly make sense to a narrower audience that are engaged in actively researching and comparing specific products or services.⁷⁶

It seems counterintuitive to assert that general search engines are cheaper and offer a broader audience in a

report that alleges one of those search engines charges higher prices as a result of an allegedly dominant position.⁷⁷ From a purely economic standpoint, why would lower prices be expected on the allegedly monopolized product if not because they compete with vertical search websites? Users of vertical search websites reveal a significant amount of information about themselves including their interest in the targeted information available on these websites. This is a fundamental advantage of advertising on specialty websites.⁷⁸ Similarly, online retailers like Amazon have user information about product browsing and purchase decisions.⁷⁹

Furthermore, the CMA Report acknowledges the presence of competition from display advertising, noting that '[t]here is some evidence that display advertising, particularly on Facebook is increasingly being used for targeting in-market conversions. This suggests that Facebook may become an increasing competitive constraint on Google in the future'.⁸⁰ This assessment is confirmed more broadly in the literature.⁸¹

Thus Google search click-and-query data are far from forming a distinct market or an essential facility. Instead, others including vertical websites and online retailers have similar information of value to commercial customers and can use their data to compete with general search engines. Given the forward-looking assessment of indispensability of EU jurisprudence on the essential facilities doctrine, classifying Google Search click-and-query data as indispensable seems inconsistent with the facts of the industry.

71 Bork and Sidak (fn 61), p 666; A.V. Lerner, 'The Role of "Big Data" in Online Competition', SSRN (26 August 2014), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780, p 46 (accessed 10 February 2021).

72 Colangelo and Maggolino (fn 45).

73 'Google AdWords Alternatives: The Best Platforms for Effective SEA', 1&1 IONOS Inc. Digital Guide (6 August 2018), available at: <https://www.ionos.com/digitalguide/online-marketing/search-engine-marketing/google-adwords-alternatives/> (accessed 10 February 2021).

74 G.A. Manne and J.D. Wright, 'Google and the Limits of Antitrust: The Case Against the Antitrust Case Against Google' (2011) 34(1) *Harvard Journal of Law & Public Policy* 26–27 ('There is actually substantial reason to doubt the propriety of a narrow market definition limited to merely online search advertising. In the first place, the general defense of the proposition is neither economic nor econometric but anecdotal: Some advertisers suggest that they do not view print and online advertising as competitors, nor search contextual, or display advertising as competitors. But other advertisers clearly see the connection, especially within the constraints of limited advertising budgets.... Even where the purpose is different between the different types of advertising – brand recognition for display ads, efforts to sell for search ads – this is merely a difference in degree. Both are forms of reducing the costs of search a la Stigler – and the relevant question is whether the difference is significant enough to render decisions in one market essentially unaffected by decisions or prices in another'); M. Lao, 'Search, Essential Facilities, and the Antitrust Duty to Deal' (2013) 11(5) *Northwestern Journal of Technology and Intellectual Property* 291–94.

75 Offline advertising may also serve as an effective competitive constraint on Google and other search engines. See e.g. Goldfarb and Tucker who showed that there is a price effect on online advertisements from changes in regulations related to offline advertisement: A. Goldfarb and C. Tucker, 'Search Engine Advertising: Channel Substitution When Pricing Ads to Context' (2011) 57(3) *Management Science* 458–70.

76 CMA Report, para 5.52.

77 CMA Report, para 5.104.

78 As Salinger and Levinson note, '[a] fundamental challenge for general search engines is to ascertain the broad intent of a search, whether it be for news, an item to buy, driving conditions, or a trip to plan. To the extent that a specialty site is gathering information from the Internet, it can draw from a smaller set of information than does a general site; and it can tailor the algorithm for ranking how useful a site is likely to be for someone issuing a particular query to the category of the search.... Finally, developers of a specialty site can custom-design its users interface to elicit specific types of information. For example, travel sites design their pages to solicit from users their origin and destination and when they want to travel'. See Salinger and Levinson (fn 61), p 37.

79 'Google AdWords Alternatives: The Best Platforms for Effective SEA' (fn 73).

80 CMA Report, para 5.55.

81 Lao (fn 74), p 294 ('While advertising on Facebook is in the form of "display ads", there is no apparent economic reason for treating display ads and search-based advertising as being in distinct, relevant markets').

4.2. Will a lack of access to Google Search click-and-query data restrain innovation?

Lack of access to Google Search click-and-query data is highly unlikely to restrain innovation in the Search Advertising Market.⁸² Improvements to search engine performance and hence an improved user experience are driven largely by improvements in proprietary algorithms, which do not necessarily rely on either user feedback or increased data access. Search engines are not constrained from improving their services and competing in the absence of increases in access to user data, let alone Google Search click-and-query data. As such, it is unlikely that competition is restrained by competitors not having access to this data.

For example, Bing researchers continue to innovate, most recently with the development of a new method based on artificial intelligence to generate large amounts of labelled training image data to classify images better and faster, thereby improving image search results.⁸³ Furthermore, the success of Baidu and Naver in Korea and China suggests that innovation is alive and well. By focusing on local content, language and related services, these rival search engines outperformed Google Search in their respective countries.⁸⁴

Online controlled experimentation, or 'A/B testing', has also been shown to be capable of yielding high quality search results without access to massive data. For example, Airbnb has used this approach for search engine optimization and, more generally, 'leans heavily on iterative experimentation for nearly every product change to ... measure effectiveness and to learn as [the company] build[s]'.⁸⁵ At Microsoft, Bing deployed a simple experiment to assess the impact of changing the way the search engine displayed ad headlines:

[The experiment showed] that the change had increased revenue by an astonishing 12% – which on an annual basis would come to more than \$100 million in the United States alone – without

hurting key user-experience metrics. It was the best revenue-generating idea in Bing's history, but until the test its value was underappreciated.⁸⁶

Recent research is consistent with the notion that scale of available data is only loosely connected with the quality of search results. Chiou and Tucker examine the benefits that search engine competitors Yahoo, Google and Bing may derive from possessing large quantities of data. The authors use policy-induced variation in guidelines surrounding the length of time that search engines can store a user's data to study how the accuracy of search results change as a result of the amount of data available. They find no empirical evidence that having less data negatively affects the quality of search results; they posit that historical data may be less useful for accurate predictions of current behaviour than is commonly believed.⁸⁷

Aridor, Che and Salz examine the effect of the EU's General Data Protection Regulation (GDPR) on a firm's ability to collect consumer data. The GDPR consent requirement allows consumers to block a website from collecting and sharing their personal data. The authors find that, while the GDPR reduced data availability, the firm's ability to predict consumer behaviour using the intermediary's proprietary machine learning algorithm did not significantly worsen as a result of the changes induced by GDPR. If anything, the firm's predictions improved, perhaps because those consumers who opt in and share their data have more complete search histories.⁸⁸ Therefore, it may be the type of data collected that matters, not the quantity.⁸⁹

Ultimately, recent academic research suggests that innovation in search engines is alive and well and that multiple alternatives to Google click-and-query data are available. These alternatives along with innovation and advancement of algorithms yield high quality search results as well as useful information for potential advertisers interested in reaching specific customers with well-defined preferences.

82 Or the digital advertising market more broadly.

83 Lee, He, Zhang and Yang (fn 67); Bing Blogs (fn 67).

84 W. Zhao and E. Tse, 'Competition in Search Engine Market' (2011) 28(2) *Journal of Business Strategies* 124.

85 B. de Luna, 'Experimentation & Measurement for Search Engine Optimization', Medium.com (25 September 2018), available at: <https://medium.com/airbnb-engineering/experimentation-measurement-for-search-engine-optimization-b64136629760> (accessed 10 February 2021).

86 R. Kohavi and S. Thomke, 'The Surprising Power of Online Experiments', *Harvard Business Review* (September–October 2017), available at: <https://hbr.org/2017/09/the-surprising-power-of-online-experiments> (accessed 10 February 2021).

87 L. Chiou and C. Tucker, 'Search Engines and Data Retention: Implications for Privacy and Antitrust', NBER Working Paper Series (September 2017), <https://www.nber.org/papers/w23815.pdf> (accessed 10 February 2021).

88 'We provide evidence that this pattern is consistent with the hypothesis that privacy-conscious consumers substitute away from less efficient

privacy protection (e.g. cookie deletion) to explicit opt out, a process that would reduce the number of artificially short consumer histories'. G. Aridor, Y.-K. Che and T. Salz, 'The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR', SSRN (29 January 2020), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3522845 (accessed 10 February 2021).

89 This is also supported in research by Schaefer, Sapi and Lorincz on data from Yahoo. The authors noted the importance of the type of data (e.g. personalized information) and the overall quality of the search algorithms on the quality of search results. The authors also found diminishing marginal returns on search quality improvement after a given terms was searched for more than 500 times. See M. Schaefer, G. Sapi and S. Lorincz, 'The Effect of Big Data on Recommendation Quality. The Example of Internet Search', DICE Discussion Paper (March 2018), available at: https://www.dice.hhu.de/fileadmin/redaktion/Fakultaeten/Wirtschaftswissenschaftliche_Fakultaet/DICE/Discussion_Paper/284_Schaefer_Sapi_Lorincz.pdf (accessed 10 February 2021).

5. Is there an objective justification for Google Search to deny access to its click-and-query data?

Our discussion in Section 4 suggests that Google Search click-and-query data are far from an essential facility. There are ample alternatives available to companies competing in the Search Advertising Market, and lack of access to Google's click-and-query data seems to coexist with robust innovation. Nevertheless, it is worth examining whether denying access to Google Search click-and-query data protects incentives to invest and innovate or protects user privacy, thereby indicating that data sharing may result in welfare losses instead of welfare gains.

5.1. Incentives to invest and innovate

The CMA Report data access remedy is intended to address the sources of market power in the Search Advertising Market. It rests on the premise that third-party search engine access to Google Search click-and-query data will allow competing search engines to improve their search algorithms, and that data access would reduce Google's economies of scale advantages over current and potential competitors.⁹⁰ Based on our findings above that such a premise rests on shaky ground, any demonstrable negative effect on the incentives to invest and innovate or on privacy must also be examined.

As the CMA Report acknowledges, it has been the quality of Google Search that has made Google more successful in attracting user attention.⁹¹ Joint demands by users and advertisers incentivize Google to continuously enhance the quality of Google Search.⁹² Users value the ability to quickly find relevant information on the internet. Advertisers value the ability to access search users with well-defined preferences. Google is in an innovation race whereby any decrease in search accuracy and quality experience for its users would also result in diminished quality of information about the searchers for advertisers. This leaves Google with significant economic incentives to innovate and attract users to its search engine by providing information relevant to their queries, quickly and in a user-friendly format.⁹³ Failure

to continuously invest and innovate would result in vertical search websites or digital marketplaces eroding Google's main source of revenue. In other words, the existence of competition described in the previous section is also the driver of continuing incentives to innovate for Google.

From a purely economic standpoint, compulsory third-party search engine access to Google Search click-and-query data is likely to have negative effects on innovation and investment. First, it will lead to lower levels of investment and innovation by actual and potential competitors to Google. As discussed in Section 4.2, access to Google's click-and-query data is not a necessary prerequisite for innovating and competing in the search market, and indeed it is possible to invest in algorithmic improvements in the absence of access to this data. However, instead of promoting creativity and greater competition in these areas, compulsory data access would allow competitors to free ride on the investments made by Google rather than investing themselves, reducing innovation in the search space.⁹⁴

In providing access to Google's click-and-query data, which are in part the outcome of Google's technological innovations and algorithmic advances, competitors can forgo investing in long-term strategic advances in their own search algorithms, instead investing just enough resources to mimic the performance of Google Search. That is, competitors can merely replicate the performance of Google Search, or adopt their algorithms to mimic its performance, rather than investing in unique and forward-looking advances and research and development projects.⁹⁵ Furthermore, the existence of compulsory data access will reduce incentives for competitors to become market leaders, thereby becoming subject to the same data access regulations. This reduction in investment by competitors reduces the level of competition in the market, reducing consumer welfare and contradicting the recognized purpose of antitrust law.⁹⁶

Secondly, compulsory access for competitors to Google's click-and-query data will directly disincentivize Google to innovate on its search engine since any improvement in search that translates into increased value of the search data for potential customers would be immediately shared

90 CMA Report, para 90.

91 CMA Report, para 5.50.

92 Bork and Sidak (fn 61), p 667.

93 Lao (fn 74), p 313; Bork and Sidak (fn 61), p 668.

94 Colangelo and Maggolino (fn 45).

95 Recent studies have shown the ability of researchers to replicate, backwards-engineer, or in effect steal machine learning algorithms where only the outcome of the algorithm is made available through an application programming interface (API). Without having direct access to the parameters of the algorithm it is possible to replicate its performance

through the provision of inputs and the collection of associated outputs to the algorithm via the API. Further, this research shows that only a small sample of queries and collection of their associated output is necessary to extract a 100% equivalent model. (See F. Tramèr, F. Zhang, A. Juels, M.K. Reiter and T. Ristenpart, *Stealing Machine Learning Models via Prediction APIs*, USENIX Security Symposium 2016 and B. Wang and N. Zhenqiang Gong, *Stealing hyperparameters in machine learning* (2018) IEEE Symposium on Security and Privacy 36–52.).

96 Bork and Sidak (fn 61), p 665.

with its competitors.⁹⁷ If competitors can utilize Google's click-and-query data to readily adapt their algorithms to match the performance of Google Search, then the incentive for Google to make continued investments in improving its algorithms is greatly reduced.⁹⁸

By disincentivizing forward-looking, long-term innovations and incentivizing free-riding by competitors, the likely result of providing data access is reduced innovation by rivals, reduced innovation by Google and reductions in ongoing improvements to online search generally, all leading to a reduction in consumer welfare.

5.2. Privacy and control over personalized data

Another important concern in the UK and across Europe is the privacy and data control of users. In particular, the GDPR gives consumers the ability 'to make an informed, specific, and unambiguous consent to the processing of their data.... [B]y denying consent, a consumer can block a website from collecting personal data and sharing it with third-party affiliates'.^{99, 100}

Despite the CMA's stated concerns about user privacy as featured in its report, the CMA states that it is 'confident that such an intervention could be designed in a way that does not involve the transfer of personal data, and hence without raising privacy concerns'.^{101, 102} But recent research has shown that it is specifically personalized information that is valuable to improving search engine performance.¹⁰³ The stated purpose of data access – to help competing search engines improve their algorithms – may only be achievable by giving access to the data that would naturally raise privacy concerns.

In other words, any sharing requirement enacted by the CMA would likely either result in data sharing of little value or an action which, may be prevented by privacy regulations.¹⁰⁴ Data protection laws stipulate that users will share their personal data with a company only if

they clearly know the specific use the company will make of it. As Colangelo and Maggiolino note, 'asking users to give a loose consent in favour of any digital platform's rival asking for the data would be contrary to the principles informing data protection laws'.¹⁰⁵

The importance of privacy considerations in data access cases is not new. In *IMS Health*, privacy considerations were one reason why the European Commission required IMS Health to grant licenses for the use of the 1860 brick structure. The creation of an alternative structure could – via a comparison against the 1860 structure – enable the identification of individual pharmacies and violate German data protection laws.¹⁰⁶ However, in contrast, given the nature of Google Search click-and-query data and their role in improving search engine performance, privacy concerns present a significant obstacle to the sharing of data if these data are to have any value to their intended beneficiaries.

6. Conclusion

The CMA Report proposed a number of remedies to ameliorate antitrust concerns and improve competition in the market for digital advertising. One remedy would require Google to give rival third-party search engines access to Google Search click-and-query data. This remedy implicitly applies the essential facilities doctrine in its assessment of the competitive conditions in the digital advertising market.

Prior EU jurisprudence on the essential facilities doctrine combined with fundamental economic observations suggest that Google Search click-and-query data do not constitute an essential facility. The data are not necessary to compete in the digital advertising market and their absence does not limit innovation. In fact, were Google to share data with third parties, investment and innovation incentives would be reduced for Google, its competitors and the broader technology community. Third-party data access would also conflict directly with privacy concerns.

97 Hylton (fn 9).

98 Colangelo and Maggiolino (fn 45).

99 Aridor, Che and Salz (fn 88).

100 The GDPR was retained in UK domestic law until the end of the Brexit transition period on 31 December 2020. GDPR has been retained in domestic UK law, however, the UK has the independence to review the framework and apply amendments. See information provided by the Information Commissioner's Office (ICO) available at: <https://ico.org.uk/for-organisations/data-protection-at-the-end-of-the-transition-period/data-protection-at-the-end-of-the-transition-period/the-gdpr/> (accessed 10 February 2021).

101 CMA Report, para 90.

102 See e.g. CMA Report, paras 37–39, 85, 5.313–5.320.

103 '[P]ersonalized information is crucial. In particular, our analysis reveals that the quality of results displayed to queries for which the search engine saw more personalized information (i.e. could track the identity of searchers for a longer sequence of previous searches) improves faster than that of queries with less personalized information available'. See Schaefer, Sapi and Lorincz (fn 89).

104 C. Tucker, 'Online Advertising and Antitrust: Network Effects, Switching Costs, and Data as an Essential Facility' (2019) 2 CPI Antitrust Chronicle 1.

105 Colangelo and Maggiolino (fn 45).

106 Evrard (fn 54), p 26.