

Why real big data may not matter that much and why data portability is crucial

Jan Kupcik¹

I. Brief note on terminology

The phrase “big data” is so often used that it sometimes looks like it might have lost its original meaning somewhere on the way. It is often generally used as a synonym to any data gathered on the internet. It is, however, questionable. Big data is “*the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value*”.² The main focus with respect to big data is to unstructured data.³ The primary element and essence of big data is in its volume and variety. Therefore we should understand data as being big only in cases where the pieces of data are grouped into something that is impossible to process by regular hardware and software tools.

Often, when turning into big data, the output created by them does not interfere with protection of personal data as it is in a form of anonymised information (describing the overall trends among a group of users, customers etc.). However, the input is very often (if not always) personal data.

In order to cover the data before becoming (in combination with other data) big, I propose to define a category of “individual data”. Such data appear in the period between the collection of “raw” data and the moment of being grouped with other data collection targets’ data into big data. It is going to be very often personal data. However, individual data may be also data not about a natural person, but about any individual user of the internet – as an example, take a social media profile of a company being managed from the same IP address by various employees.

II. Reason for another category of data

The differentiation of big data and individual data is not self-standing in my analysis. Individual data only may be used for targeted advertising. Big data would have to be decomposed to individual data before applying targeting. Individual data can provide “personalised” or rather “individualised” search results and improve other digital services by means of individualisation.

First of all, individual data are vital for online advertising in its current state. Not many advertisement providers may survive without targeting as it would be ineffective and costly - which is exactly the reason why offline advertising share of overall advertising market constantly decreases over the years. Should targeting remain allowed - and it should, as it is simply more effective way of advertising that will, in a long-term, result in efficiencies shared with consumers – the online advertising providers’ focus is going to be on individual data. Only in the second step, it is big data coming to spotlight.

¹ The author is a Ph.D. candidate at Masaryk University in Brno, Czech Republic. He is also associated with Prague office of Weil, Gotshal & Manges.

² De Mauro, Andrea, Greco, Marco and Michele Grimaldi. A Formal definition of Big Data based on its essential Features. Library Review. 2016 (65), Vol. 3, pp.122-135.

³ Dedic, Nedin and Claire Stanier. Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery. In: Piazzolo F., Geist V., Brehm L., Schmidt R. (eds) Innovations in Enterprise Information Systems Management and Engineering. ERP Future 2016. 2017, Vol. 285. Springer.

In a case of online advertising, it is apparent that individual data are core for provision of the services. I admit it is not entirely the same in the area of genuine search. However, I believe that most people would be surprised how their search queries look like if they were deprived of the benefit of individualised search results.

What I am trying to argue is that we cannot prematurely consider data as being homogenous group and we should, before attempting to find competition law solutions, analyse and classify data. Only after that we can discuss various competition law solutions with respect to data in digital economy – as, for example, potential application of essential facility doctrine and duty to share data.

III. Prematurity of essential facilities doctrine

It has been often suggested to use competition law as a tool for creating a data sharing obligation of certain companies, which allegedly should foster competition. I believe it would not be lawful, but even more, it would not be legitimate and appropriate to make certain companies to share personal data with any of its competitors.

First of all, we cannot consider data as one category. In rare cases of refusals to deal the undertaking in a dominant position should be obliged to deal in the narrowest possible sense. Hence, data should be divided into different categories and only the really indispensable part of data should be shared. This is the reason why I believe that differentiation of big data and individual data is necessary. At least in the segment of online advertising, it is the individual data that is necessary to target advertising and, therefore, to operate on the market.

In many cases, individual data will be personal data as well. Privacy, including personal data, may be decreasing with the rise of the Internet. Nevertheless, people cannot be deprived of privacy in total as it is an essential feature of democracy. They may want to share certain data with certain subjects, but never all data with a subject or some data with all subjects. Personal data sharing obligation must be avoided if any viable option is present.

IV. Data portability

For that matter, I am of the opinion that data portability principle (as currently in a narrow version provided in GDPR) should be analysed and applied. Mutual relationship of data portability and competition law has already been studied.⁴ I believe that expanded data portability obligation could even forces in at least some of the digital markets (including online advertising segment), as opposed to tipping the balance to the hands of often less efficient competitors by imposing a data sharing obligation via essential facility doctrine.

Data portability principle would have to be applied not only to volunteered data, but also to observed and inferred data. Competition law should provide other safeguards against

⁴ See for example Inge Graef, Jeroen Verschaleken, Peggy Valcke, 'Putting the right to data portability into a competition law perspective' (2013) *Law: The Journal of the Higher School of Economics, Annual Review* 4. Available at SSRN: <http://ssrn.com/abstract=2416537>; Diker Vanberg, A. & , Ünver, MB., "The right to data portability in the GDPR and EU competition law: odd couple or dynamic duo?", in *European Journal of Law and Technology*, Vol 8, No 1, 2017; Graef, Inge 'Data portability at the crossroads of data protection and competition policy' *Big Data e Concorrenza* - organised by the Autorità Garante della Concorrenza e del Mercato and LUISS Osservatorio di Proprietà intellettuale Concorrenza e Comunicazioni, Date: 2016/11/09 - 2016/11/09, Location: Rome; Geradin, Damien and Kuschewsky, Monika, *Competition Law and Personal Data: Preliminary Thoughts on a Complex Issue* (February 12, 2013). Available at SSRN: <https://ssrn.com/abstract=2216088>.

consumers' lock-in as well. This setup should lower switching costs as far as to zero. In these circumstances, it will be for the competitors to fight for individual data by offering better services on the merits. Obviously, the current dominant competitors may be demotivated to a certain extent from collecting individual data as they might do the service for other competitors, which may obtain the data via data portability if a user decides to switch. However, it is still less demotivating than a wide obligation to share the data with virtually any competitor who asks for it (under essential facility doctrine). Moreover, it should also be driving the dominant competitor to invest into innovation in order to keep providing the best service (and preserve valuable data from being shared with others).

V. Essential facility doctrine could be applicable for big data

Only after thoroughly analysing the group of individual data and effect of wider data portability principle to competition, then it would be logical to focus on big data. If even after application of data portability should big data be regarded as indispensable (essential), then we can discuss the question of potential application of essential facility doctrine. This, in my estimation, will not be a case for online advertising. But it very well may be the case for mapping services. Nevertheless, this remains to be seen after detailed economic analysis and it will vary for each digital market.

I believe it could be relatively easy to construct a sharing obligation with respect to big data in a way that it would not interfere with personal data protection. For example, the dominant competitor may be required to share the data in a form of anonymised output (as discussed above).

VI. Conclusion

Discussing essential facility doctrine with respect to (all) data is premature. Its application would either (i) seriously interfere with personal data protection or (ii) would lead to requirement of anonymization of the data, which, in turn, would be ineffective as anonymised data cannot be used for targeting / individualisation. As many digital markets, including the most important one⁵ – online advertising market, are primarily based on individualisation, possession of individual data may be decisive. However, this may be solved by data portability principle, if widened to all kinds of individual data (volunteered, observed and inferred). It is less interfering with legitimate interests of the dominant players as it only causes sharing of data with others in cases when others win the user on the merits (through a better service). It creates an additional incentive to the dominant players to innovate and focus on better services, rather than how to lock-in the users or leverage their dominant position somewhere else.

Consumers are also protected from having their personal data shared with third parties, unless they decide to switch. I believe that consumers would be well aware that the decision to switch is in the same time going to be understood as a consent to provide their data to another service provider.

Only in this perspective we should analyse whether big data are indispensable as well. I believe that in many cases it will be shown that big data are not that important after all.

⁵ Compared to other areas of digital economy, the revenues from advertising seem to be the most significant.