Innovative Policy Instruments for Effective Market Oversight[1]

Dr. Nicola Jentzsch (Data Economy Lead)

- Please only publish at acceptance -

I. Introduction

Digital markets are increasingly dominated by companies that have built up large-scale ecosystems of products and services that stretch across different markets. In these ecosystems data are accumulated by Tera- and Petabytes. At the moment, the implications of such ‘data power’ for competition are unclear and there is a controversial debate over how to reform Competition laws.

The main goal of this paper is to propose an innovative set of instruments for market oversight. The proposed tools are intended to effectively limit market power and to set incentive schemes that drive competition towards value- and privacy-sensitive innovation in future. The latter is important for GDPR-compliance and for cyber-security. Supervisors in charge for competition oversight, but also Data Protection Authorities may deploy these instruments for different use cases in a harmonized way across Europe. Moreover, these instruments can be implemented at scale. They are also welfare-enhancing as they constitute binding commitments for companies, e.g. in merger cases, which reduces legal uncertainty. Moreover, they allow companies to commit to a specific level of information-sharing.[2] IT-giants like Google, Amazon, Facebook and Microsoft merge data across their portfolio. The quality of predictions of models deployed in different services and the latency of queries are the main differentiation parameter in competition for marketing, medical or

---

[1] This paper is an excerpt of the following publication: Jentzsch, N. (2018). Marktmacht in der Datenökonomie begrenzen - Aktuelle Herausforderungen der Wettbewerbsaufsicht in der Digitalwirtschaft, Impulse (August 2018), https://www.stiftung-nv.de/de/publikation/marktmacht-der-datenoekonomie-begrenzen

[2] Akçura, T.M. and Srinivasan, K. (2005). Customer Intimacy and Cross-Selling Strategy, *Management Science* 51: 1007–1012.

mobility services. Thus, companies are subject to quality competition. In such fast markets, where data flows and models constantly change we need to develop new tools to keep supervision effective.

In the following, the (problematic) scale and scope effects deriving from Big Data architectures are discussed. In this part, we analyze bottlenecks such as non-replicable datasets and patents on machine learning methods.[3] We discuss how companies – through exclusionary licensing behavior – create substantial entry barriers for third-party providers for analytical tools, for example. This part addresses the question concerning problematic data exchange and problematic data pooling, emphasized in the Commission's call.

Second, we analyze the dynamics of the large-scale mergers witnessed in the past years. It is discussed how the Commission could have bound companies to their promises to not deteriorate their data protection policies pre-/post-merger. The strategic data compliance actions after Google/DoubleClick (2007), Facebook/Whats App (2014) und Microsoft/LinkedIn (2016) created an information oligopoly, where each company is differentiated from the other, yet overall the *combination of datasets* in firms is not replicable. We introduce an instrument by which such non-replicable datasets could be shared with potential competitors, while at the same time preserving the privacy rights of data subjects. This part directly addresses the question how we can ensure AI technology becomes as competitive as possible.

Finally, we discuss how privacy aspects (*not concerns*!) should be incorporated into the analysis of company behavior either in merger cases or cases of misuse of market dominance. We conclude by discussing the pros and cons of the proposed instruments.

II. Scale and Scope Effects in Digital Markets

One of the most controversially discussed questions is whether increasing returns to scale exist in the usage of Big Data, which creates 'escape competition.' Some authors refute this

[3] An efficient organization of databases is as well important, but we cannot elaborate on this matter.

claim,[4] others seem to be in slightly in favor,[5] whereas a third group of scholars quantitatively proves increasing returns to scale in individual use cases such as eCommerce.[6] At this stage it is clear that it depends on the individual case, whether scale or scope effects arise through the usage of Big Data.

2.1 Sparseness drives Preference for Variety

Sparseness arises in Big Data environments where the number of interesting events – such as purchase transactions by individuals on the Internet – is very low given the size of the data matrix. If sparseness is given, the prediction quality of different models increases with the amount of available data. This occurs at a decreasing rate, though. However, it drives companies to incorporate an ever greater *variety* of data types into their portfolio. Adding different data types, collected from IoT devices, fitness trackers, mobile phones and desktop computers, adds value to the existing data portfolio and improves quality of predictions.

In an environment, where data flows, models and use cases constantly change it is virtually impossible to understand whether adding a new data source to a portfolio (e.g. by merger) evokes increasing returns or not. That aside the drive for variety clearly has privacy implications, in particular, if the data is extracted from individual users, who cannot really join another platform.

2.2 Replicable and non-replicable Datasets

While some datasets are clearly replicable, also at speed, others just aren't. The latter is in particular the case for data generated by company-specific applications. Likewise, in cases where there are exclusionary patents on applications/methods that generate data, data sets might not be easily replicable either.[7]

IT-companies in general *neither license nor sell* datasets themselves. Competitors *do not have access* to their data and cannot easily re-create, say, the Google data portfolio.

---

[4] Varian, H. (2017). Artifical Intelligence, Economics, and Industrial Organization, https://www.nber.org/chapters/c14017 und Chiou, L. und C. Tucker (2014). Search Engines and Data Retention: Implications for Privacy and Antitrust, MIT Sloan School Working Paper 5094-14, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2441333.

[5] Junqué de Fortuny, E. J., D. Martens und F. Provost (2013). Predictive Modeling with Big Data: Is Bigger Really Better? Big Data 1 (4): 215 – 226.

[6] De Cnudde, S., D. Martens, T. Evgeniou und F. Provost (2017). A Benchmarking Study of Classification Techniques for Behavioral Data, Dept. of Engineering (University of Antwerp) Research Paper 2017-005 (April).

[7] Licensing fees can be fairly high, which would put a competitor who tries to use the same technology to generate data on unequal footing.

Competitors would have to enter several markets with freemium services in order to rival a given dominant player's data collection. Through adding new information sources, a dominant player's data collection becomes *more unique* and achieves greater 'information depth' per capita (of data subjects that is).

At the moment, the Commission neither has metrics on size, variety or depth of datasets accumulated in the dominant market players that could improve on transparency about what is going on – information-wise – and guide its decisions.

2.3 How to Exclude Competitors: "Drive-by Licensing"

The competitive strategy of firms is to establish a broad tech ecosystem first and then to 'milk' those that have become depend on the system. Once the ecosystem thrives, licensing can be used to drive out competitors or to control potential competitors. This is what we – in a somewhat exaggerated way – call 'drive-by licensing'. The goal is to create closed licensing systems. For example, if an automobile company hosts data in a cloud platform, the platform can start licensing third-party access to the data of the automobile company. These licensing fees can be *much higher* compared to the fees for tools offered by the hosting service itself. This leads to a lock-in effect of the stored data and reduces the chances of third parties to provide competitive services.

Licensing also goes together with patents. This is an additional point worth mentioning here. Representatives of large IT-companies often characterize patents as not being important. In fast markets they quickly become obsolete, it is argued. However, globally we witness an *increase* in patent applications on machine learning methods especially in the US, China and South Korea. Companies like Microsoft, IBM and Google are leading the statistics of patent applications in this area.[8]

We conclude this section with three main insights:

- Depending on the use case, there can be decreasing, increasing and constant returns to scale arising due to deployment of Big Data technologies;
- There are replicable datasets, but there are also non-replicable ones generated by company-specific applications in the data eco-system of a firm; and

[8] Döbel, I., M Leis, D. Neustroev, H. Petzka, A. Riemer, S. Rüping, A. Voss, M. Wegele und J. Welz (2018). Maschinelles Lernen – Eine Analyse zu Kompetenzen, Forschung und Anwendung, Fraunhofer-Gesellschaft, S. 20 – 21.

- Once an ecosystem has been established, companies can use 'drive-by licensing' to control competition by third-party providers of tools.

III. Creating Ecosystems: Mergers in Digital Markets

Three merger cases have attracted special attention in the past because of their sheer size and strategic importance for the markets' future development trajectory: Google/ DoubleClick (2007), Facebook/WhatsApp (2014) und Microsoft/LinkedIn (2016).[9] In all cases, it can be shown that companies strategically used their data protection compliance policies to create unique data portfolios.

After the merger in 2007, Google initiated a complex succession of data policy changes that resulted in the global merger of virtually all of its databases (see Global Data Policy of 2012).[10] This by far exceeded the envisioned merger of only two systems, Google search behavior and DoubleClick webtracking of users. Its potentially exclusionary usage of its marketing stack is now under supervision by the German competition authority.

In 2014, Facebook purchased the much larger messenger service WhatsApp. This service had about 600 million users globally, far more than the 250 million of Facebook's messenger. The Commission saw no negative effects for competition arising from the merger. However, competition here follows the 'rules of user attention:' users that chat on WhatsApp cannot simultaneously chat on FB messenger.

Facebook was looking for a way to bind younger users early to its platform in order to strengthen its core service. WhatsApp, on the other hand, promised that privacy terms would not change for its users. About two years later Facebook tried to exchange phone numbers and user metrics with WhatsApp.[11] Since then, Facebook has completely 'assimilated' the service: It stopped the paid-service policy in 2016, introduced encryption

---

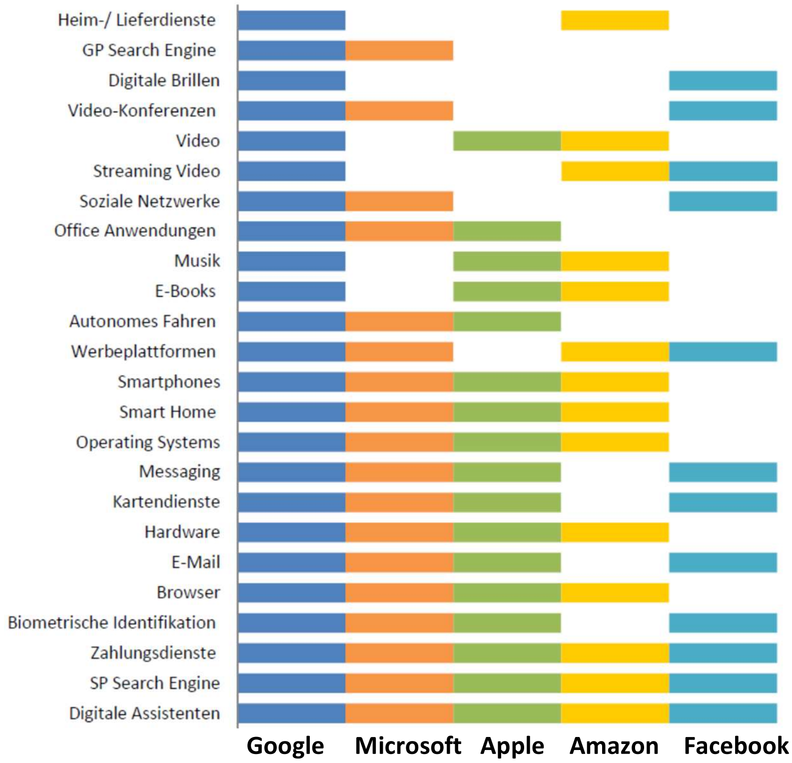
[9] For details, the reader is referred to the Commissions decisions as well as Jentzsch, N. (2018). Marktmacht in der Datenökonomie begrenzen - Aktuelle Herausforderungen der Wettbewerbsaufsicht in der Digitalwirtschaft, Impulse (August 2018), https://www.stiftung-nv.de/de/publikation/marktmacht-der-datenoekonomie-begrenzen

[10] Dutch Data Protection Authority (2013). Investigation into the combining of personal data by Google – Report of Definitive Findings, November, z2013-00194, https://autoriteitperso-onsgegevens.nl/sites/default/files/downloads/mijn_privacy/en_rap_2013-google-privacy- policy.pdf.

[11] This was punished by the Commission with a fine of about 110 million Euros.

(based upon WhatsApp patents) in 2016 and announced in 2018 that WhatsApp from now on will be 'supported' by advertisement.

Fig. 1 Portfolio of Different Services and Product



Facebook and WhatsApp were differentiated by *privacy quality*. Therefore, this merger led to the disappearance of a privacy-friendly alternative from the market. WhatsApp had put *competitive pressure* on Facebook by virtue of its business model. What if WhatsApp would have founded a privacy-friendly social network? As a negative side effect the purchase of WhatsApp implies that other privacy-friendly alternatives on the market are not trustworthy anymore. After all who guarantees the users that those services are not also bought up one of the major players? In information markets, differentiation exerts competitive pressure.

In 2016, Microsoft made a bid for LinkedIn, the largest professional social network on the globe. Again, the Commission held that the companies did not qualify as competitors. This time, the Commission allowed the merger with conditions, but none of these had *data practices* as a subject. The Commission refuted any concerns with respect to 'data power' by stating that a large share of data would be available to the market post-merger. This is an incomplete view of the matter and could now even be qualified as wrong. It is incomplete, because it neglects any 'ecosystem' data portfolio effects. It is wrong, because post-merger Microsoft obtain access to far more information compared to competitors.

With the merger, Microsoft obtained complete access to the dynamic LinkedIn data (including meta data, social ties, etc.), and not only static 'Business Card' information, which was typically accessible for advertisers. The merger created an information asymmetry that did not exist in the market before. It enabled Microsoft to merge data from LinkedIn with many other sources (e.g., Windows-10 usage). Commentators saw now problem here as Microsoft was not driven by advertising revenues. After two years, however, Microsoft announced the establishment of its Microsoft Audience Network. For this Network it merges data from Bing, MSN, Outlook, Skype and LinkedIn, which are analyzed by Artificial Intelligence for targeted advertising. Again, the Commission did not flank this merger with binding commitments with respect to the data practices. We conclude with three main insights:

- Comparing individual services and products wholly falls short of understanding the company ecosystem and 'data portfolio effects';
- Today's mergers involve several databases to create unique user graphs for targeting or other services; and
- There are currently no analytics, metrics and commitments that would counteract these developments.

IV. New Theories of (Privacy) Harm

At the moment, we do not have theories of harm that incorporate privacy. The reason might be that privacy itself is a fuzzy concept. In this context, privacy denotes an asymmetric distribution of private and personal data.[12] 'Privacy harms' refer to a change in the distribution to the detriment of the data subjects. Such a change results in tangible (price-related) and intangible (non-price-related) welfare effects.

Robust theories of harm are absolutely necessary for making a case that survives a judge's scrutiny in court. We start by differentiating material and immaterial privacy harms. Note at the outset that harms cannot always be traded-off against benefits, in particular if the harm is a violation of a fundamental right. Note also that we do not speak about 'privacy

[12] 'Private' denotes that the information is not common knowledge. The term 'personal' denotes the differentiation power of the data, i.e. whether it is useful for identifying persons.

concerns.' These are an individual's perceptions of risks, which we do not see as part of the competition analysis.

- **Material privacy harms** – Database mergers result in an improved information base for customer segmentation.[13] Segmentation may involve price discrimination[14] or product personalization (or both). Segmentation induces welfare effects. These are negative if the rent extraction effect out-weights the competition effect.[15] In such competition, consumer rents are shifted away and their welfare is reduced. Moreover, personalization can increase switching costs for consumers. Welfare is also reduced due to a smaller choice set. Post-merger a *relevant* (anonymous) option has disappeared from the market.
- **Immaterial privacy harms –** Immaterial harms that arise from increased information sharing and pooling are very hard if not impossible to quantify. Another harm is that individuals may become subject to manipulation depriving them of individual autonomy. This is the case if data provided for social networking is, for example, used for voter manipulation. Consumers could also be directed to 'grow into' using dominant platforms, which capture them through network effects. Given some path dependency, they will forego the chance to grow into another (better) platform that is more privacy-friendly.

V. Future Instruments of Effective Market Oversight

Oversight authorities need to employ different instruments to remain at par with tech companies in fast moving markets. In the following, we will discuss some of them and note their advantages and disadvantages.

Understanding ecosystems: Mergers create datasets that are become more unique by the day. Merged datasets can be used to improve core services. Authorities could better understand companies by creating digital twins, which represent the ecosystems or

[13] The incremental improvement by adding information items can be quantified. An example is provided in Winters, P. (2018). KNIME in Action: GDPR, Taking a Proactive Approach, Presentation at the KNIME Spring Summit.
[14] Privacy harms can in fact be quantified. For example, the difference between a
[15] Ghose, A., K. W. Huang (2009) Personalized Pricing and Quality Customization, Journal of Economics & Management Strategy, 18. (4): 1095-1135.

particular business models. This could create an overview of the ecosystems deployed by firms.

Privacy Guarantees: Mergers with potentially negative effects for data protection ought to be subjected to formal privacy guarantees. Formal privacy guarantees such as Differential Privacy, k-anonymity are routines that can be used by developers. Their main advantage is that they are provable and they can be dynamically adjusted. In general, there is a trade-off between data utility and privacy. A firm in a concentrated market, for example, would not be allowed to merge data at the individual level, but only at a specific aggregated level, leaving (quality) space for competitors to compete.

Privacy guarantees are aligned with GDPR-goals and can be shared across Europe. Moreover, these methods can be vetted by *independent experts*.[16] A logging of these also allows the DPAs to keep track of the 'state of the art' of depersonalization.

Data Synthetization and Pooling: Thus far, the sharing of data amongst competitors was refuted for data protection purposes. However, once datasets are not replicable, it is extremely difficult for competitors to enter markets. Data synthetization allows to create an 'artificial' representation of original data. Unlike other techniques it preserves important associations between features. The sharing of synthetic data, however, *must* be accompanied by strict purpose definitions, otherwise it undermines data protection regulations. Purpose definitions can be implemented at scale in automated decision systems.[17] Training of model on synthetic data reaches almost the same quality as on original data.[18]

Experimental Evidence: Experimental evidence, in particular the evidence from the economics of privacy, ought to be much better included in analyzing competition cases. Just because user *could* switch between services does not mean that they in fact do so. Many services put a burden on customers if they want to deport data and there are no no-click solutions to erase a profile. Also, data portability must be seen critically from the point of

---

[16] An example is the critique by several experts of Apple's application of Differential Privacy, see Becker, L. (2017). Wissenschaftler: Apples Differential Privacy ist löchrig, https://www.heise.de/mac-and-i/meldung/Wissenschaftler-Apples-Differential-Privacy-ist-loechrig-3843221.html?wt_mc=rss.ho.beitrag.rdf

[17] Lizee, Z. (2018). Formalizing Use Restrictions in Privacy and Data Privacy Agreements, Presentation, Harvard De-Id Group.

[18] Apple (2017). Improving the Realism of Synthetic Images, Apple Machine Learning Journal 1 (1) https://machinelearning.apple.com/2017/07/07/GAN.html

view of customer mobility. The emphasis should be lee on possibilities and more on real behavior in markets.

Deployment of Machine Learning: In order to track major indices in fast markets, supervisors also need to deploy machine learning tool. There are a number of open source platforms that can be used for such tracking. Price movements aside, authorities could also detect collusive behavior in terms of data collection practices. US- researchers show how data protection guidelines can be automatically loaded, analyzed and tracked across hundreds of companies.[19] This could be done by the Commission or it could be outsourced to an independent third-party. At this stage, it is unclear how authorities – with the tools they deployed in the past – want to remain effective in future. The current tools as well as very narrow timelines for the analysis of merger cases render competition policy virtually ineffective.

VI. Conclusions

In today's digital markets, competitors differentiate via quality of their products and services, in particular in 'freemium markets.' The quality of their services is based upon social graphs that become more unique by the day, because of the accumulation of personalized and depersonalized data. With the development of product and price personalization – virtually the destined future of all smart services – competition and data protection questions converge. Companies produce the social graphs by creating ecosystems, which give rise to complex competitive dynamics. Even economic theory often reaches its limits in explanatory power in such complex settings.

In order to deal with such development, competition policymakers must develop new instruments. This paper proposes the transfer of some instruments from their application in data protection to competition policy. The goal is to find new instruments for a more effective and more systematic oversight. Even if the proposed tools are not used in their 'pure form', combinations therefore could prove to be a way to keep at par with tech companies in fast moving markets.

[19] Cranor, L.F., K. Idouchi, P. Giovanni, M. Sleeper und B. Ur (2013). Are They Actually Any Different? Comparing Thousands of Financial Institutions' Privacy Practices, WEIS 2013, https:// www.blaseur.com/papers/financial-final.pdf.