Competition and Data Protection Policies in the Era of Big Data: Privacy Guarantees as Policy Tools

Dr. Nicola Jentzsch*

Abstract

Privacy guarantees hold considerable potential for expanding the toolbox of instruments of Competition and Data Protection Authorities and for improving the effectiveness of supervision. This paper introduces to concepts such as anonymity (identifiability), pseudonymity, depersonalization, differential privacy, among other concepts. For Data Protection Authorities, it is discussed how these concepts can be used to determine the scope of supervision. Moreover, I elaborate on how Competition Authorities can use these concepts in merger cases. The paper is intended to bridge the existing gap between academic concepts and practical policy-making. It suggests a more rigorous approach to supervision in order to deal more effectively with firms that use Big Data and personalization in competition.

JEL-Classification: *L40; L14; D18.*

Keywords: *Competition, privacy, data protection*.

* Corresponding author: Deutsches Institut für Wirtschaftsforschung (DIW Berlin), Mohrenstr. 58, 10117 Berlin, Germany, T. +49-(0)30-897-89-0, Fax +49-897-89-200, njentzsch@diw.de. First draft, comments are welcome.

This paper reflects the opinion of the author and is not a public statement of DIW Berlin or its associated institutions.

I Introduction

Information asymmetries are a key ingredient for competition. Between rival firms, they ensure the protection of a company's valuable trade secrets and induce uncertainty about the competitor's next movements or innovations. Between firms and consumers, information asymmetries uphold informational rents derived from an economic transaction. Moreover, they imply uncertainty about a customer's switching intentions. This induces market discipline in competition – at least to a certain extent – as it is uncertain under what conditions a firm's customers might migrate to the competitor, providing an incentive to serve them better.

The increasing production and analysis of massive amounts of structured and unstructured data sets (so-called Big Data), however, has the potential to change this picture. Big Data tilts the information asymmetries that existed in the past, not only between competing firms and firms and consumers, but also between firms and supervisors. This poses a major challenge for the latter.

There are three major trends observable: First, an increasing number of firms employ methods such as machine learning and data mining for analyzing Big Data. Secondly, the analyses allow that more and more goods and services become personalized (smart cars, smart energy, and smart medicine, to name but a few). Same will be observable soon in the area of pricing, as an increasing number of firms employ personalized pricing strategies.

The increasing complexity of the analytical methods used in firms creates transparency challenges. Firms will be able to monitor consumer and rivals much better than in the past. If open markets, competitive pressure and consumer protection as policy targets are to be furthered, we need to discuss how some of the recently developed privacy guarantees can be utilized as tools for upholding information asymmetries needed to ensure competition.

The focus here is the deployment of analytical methods upholding information asymmetries in practical policy-making for supervisory authorities. The biggest promise of the new methods is their measurability. It is now formalized *how 'personal or identifying'* data traces are and in addition, we can formally state *how much information is lost*, once de-personalization methods are applied.

This paper is structured as follows: First concepts such as anonymity (and identifiability), pseudonymity, de-personalization, differential privacy and multi-party computing are introduced and explained in a way geared towards a general audience. Next I elaborate

on how Data Protection Authorities (DPAs) and Competition Authorities (CAs) can use the aforementioned concepts. Certain matters of importance, however, have to be excluded for brevity. These are aspects related to the types of attacks, which some observers say are merely a *theoretical* possibility. This plays a role, when the robustness of privacy guarantees is discussed.

The new methods hold considerable potential for improving supervisory policies as they allow DPAs and CAs to employ them as additional tools for supervision. Oversight will fail, if we do not start to find much better methods and procedures in supervising institutions that employ Big Data technologies and analytics.

II Privacy Guarantees as Policy Tools

The term 'privacy guarantee' is used herein to subsume a variety of concepts that create, reduce or uphold information asymmetries with respect to personal information compiled in databases of firms. The terminology varies and some definitions are subject of dispute. A selection of definitions is presented in the Annex. In general, there are the following families of techniques: (1) Pertubative methods such as randomization (including noise addition, permutation or differential privacy) and clustering; and (2) non-pertubative methods such as generalization (including aggregation and k-anonymity) as well as pseudonymization, see Article 29 Data Protection Working Party (2014).[1] Each of these have strengths and weaknesses with respect to risks related to singling out, linkability and inference. Some guarantees, e.g. differential privacy, are only useful, when global population statistics are computed and are less useful when the intended use is extracting information on individuals (MIT 2014: 39).

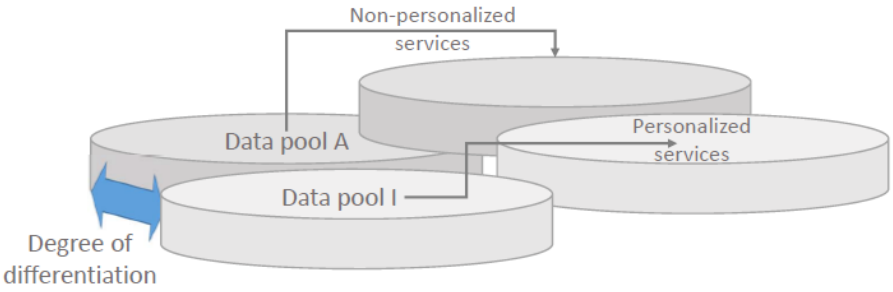
The anonymity (or privacy) provided by a system is stated in the guarantee mechanisms a firm deploys in data management, IT security policies set aside for this context. The guarantees relate to data quality and quality of queries on databases. Thus, the concepts relate (1) to data management differences between firms; and to (2) the data quality in their databases. Consider two firms, A and I, both compiling data on individuals. Firm A states it anonymizes data and provides non-personalized services, whereas firm I collects personal information and provides personalized services (see Fig. 1).

There is a degree of differentiation between both firms, which can be captured by the spectrum of identifiability of subjects in the databases of each firm. Whereas pool A

[1] There are now variants of k-anonymity to reduce inference attack risks, including l-diversity and t-closeness.

represents an anonymity set, where subjects cannot be distinguished from each other,[2] pool I hosts subjects that are identified.

Fig. 1 Differentiation Degree in De-personalization



There are four concepts mentioned here that denote the 'degree of differentiation' on the horizontal identification (anonymity) spectrum: depersonalization (anonymization), pseudonymity and identifiability or uniqueness.[3] Such a differentiation degree can also be found in competing firms that personalize services, where one firm applies more privacy-friendly techniques than the other.

It is well-known that DPAs can apply these concepts for a determination of the scope of supervision. This will be explained in greater detail below. Authorities in charge for enforcing legislation relating to unfair commercial practices can use the 'degree of differentiation' spectrum to prosecute any misleading promises of firms regarding anonymization of data.[4]

For CAs, differentiation degrees could be important for several reasons. For example, in merger cases, authorities need to define the relevant market (product-wise, geographic and temporal), before assessing dominance and its anticompetitive effects. If a merger creates or strengthens a dominant position stifling competition, it might be prohibited. Databases play a critical role in the merger of data-intensive firms or in evaluating the abuse of a dominant position. Again, this will be discussed in greater detail below. For ease of overview, the different responsibilities of the authorities are summarized in Table 1.

[2] Let us assume for now that it is possible to achieve robust anonymity.

[3] Uniqueness describes how 'identifying' a data trace is within a given set of data, i.e., how many data points are needed to achieve uniqueness.

[4] Falling short of privacy promises has already been subject of contention in a number of cases, e.g. in Federal Trade Comm'n v. Wyndham Worldwide Corp., Case No. 14-3514 (3rd Circuit, Aug. 24, 2015).

Table 1 Overview of Legislative Frameworks: United States and Germany

	Competition	Data Protection	Unfair Commercial Practices Act
Germany	Bundeskartellamt	Data Protection Authorities	Market participants (Gesetz gegen den unlauteren Wettbewerb)
U.S.	Federal Trade Commission	Federal Trade Commission (for the FCR Act)*	Federal Trade Commission (FTC Act, sect. 5)*
Goal	Protection of competition	Protection of personal data and individual privacy rights	Protection of consumers from unfair commercial practices
Activity fields	Horizontal arrangements	Data protection	Unfair and deceptive acts and practices
	Vertical arrangements		Trade secrecy leakage
	Abuse of dominant position		Pyramid schemes
	Price discrimination		Misuse of templates
	Merger Control		

* Note: FCR Act denotes Federal Credit Reporting Act, FTC Act the Federal Trade Commission Act.

2.1 Privacy Guarantees as Policy Tools for Data Protection Authorities

Reviewing the data management practices in firms is part of the DPAs daily business. Once a DPA conducts an audit, the company under scrutiny presents it with the de-personalization techniques deployed. In order to guide firms, several DPAs have published guidelines on anonymization methods (see for an example, UK Information Commissioner 2012). Herein, it is proposed to take this approach further as a combination of methods of depersonalization (anonymization) and identifiability metrics will render a more comprehensive picture about a firms' data procedures. While the first guarantees a specific level of privacy, *identifiability metrics* should be applied to evaluate the *result* of such a mechanism. Note that 'identifiability' links back to the legal term used in the definition of 'personal information' (Article 29 Data Protection Working Party 2007: 12). It is a threshold condition to determine whether information falls under the definition of personal information.

Depersonalization and the Scope of Law

It is by now empirically well established that it is very hard to achieve robust anonymization (de Monjoye et al. 2013; de Montjoye et al. 2015, de Mulder et al. 2008; Narayanan and Shmatikov 2008). For DPAs the new concepts have the following implications: first, it is now possible to formally quantify the identification degree of

information and secondly, it is possible to define that sets of a particular identification degree come under supervision.

Privacy Guarantees for On- and Off-side Supervision

While more and more firms use Big Data analytics, DPAs are chronically understaffed and under-funded. Therefore, it is worthwhile to explore how privacy guarantees can be used for off-site and risk-based supervision, for a more effective monitoring of industry players and as commitment device. The proposal is *not* meant to be a substitute for the current supervisory practices, but rather technical support to organize on- and off-site supervision in DPAs more effectively. A side-effect is that DPA and CAs could develop a harmonized way of interacting in cases of competition. The proposed approach has three main pillars:

(1) More rigorous deployment of privacy metrics related to privacy guarantees;
(2) A reporting mechanism for (selected) companies; and
(3) Analysis and integration of statistics into practical work of supervisory authorities.

First, metrics ought to be employed far more rigorously than in the past. An authority is now in the situation, where it can measure the *identification power of data*. It can quantitatively *evaluate the quality of the outcome* of the deployed de-personalization techniques in a firm, which might be often combinations of methods. Measurement allows the tracking of metrics. The basis are these two related considerations: (i) Data sets that are robustly anonymized are not under the scope of the Data Protection Law and thus the supervision of the DPA;[5] and (ii) data sets of a particular identifiability degree will come under the scope of the law and supervision.

In order to deploy metrics, the authority may either analyze the entire dataset of a given firm or – more realistically – samples of the data. Moreover, the approach should be employed for the total of datasets available and potentially linkable within a firm.[6]

Second, while evaluating privacy guarantees in firms is part of a DPAs daily business, a metrics-based approach allows the setup of a monitoring mechanism that dynamically provides metrics over an application programming interface (API). This has to be

[5] This can be qualified at times, see Article 29 Data Protection Working Party (2014).

[6] For example, one energy provider stated on a conference in Berlin recently that it 'discovered' about 80-90 different databases on customers. Until that respective research, there was no unified view on how many customer databases existed within the company.

mindful of any security matters of such reporting. If policymakers opt for a reporting mechanism, they would have to set up reporting thresholds as done in credit reporting systems (see also Jentzsch 2007) as well as standardized reporting formats. Notification thresholds could be defined by a combination of criteria such as company revenues, number of users, and/or the identifiability degree (depersonalization degree) of the data. It might also only capture the most relevant firms in a specific industry or specific cases such as in merger procedures (discussed below).

Thirdly, such reporting enables the monitoring of firms that start to compile large amounts of user data. Potentially, DPAs can fine-tune policy advice by adjusting degrees of identifiability. DPAs could track in real time changes in firms or the industry. This can support and strengthen their supervisory capacity and develop it into a more systematic instead of the currently used ad hoc approach. For example, if identifiability of data strongly increases in a firm, DPAs could opt for reviewing data collection practices and promises made towards the consumer.

There are a number of open issues. First, a reporting standard would have to be defined, especially if firms use combinations of procedures. Secondly, we have no comparative metric for the *breadth of data* compiled on individuals, i.e., how many attributes on individuals are collected. This is an important void: There is a major trend towards compiling an *increasing variety of different variables (labels) on individuals* such as behavioral and biometric data.[7] Moreover, there are metrics on 'linkability' of data sets, but not on sensitivity of data, which is often context-specific. A well-defined body of metrics may be used by DPAs and CAs alike.

2.2 Privacy Guarantees as Policy Tools for Competition Authorities

Competition in digital markets represents a major challenge for traditional competition oversight (Jentzsch 2016). Digitalization affects all fields of oversight, including horizontal and vertical arrangements, abuses of dominant position and merger control (see also Autorité de la concurrence and Bundeskartellamt 2016; Competition and Markets Authority 2015). In the following, I focus on merger control and on reaching/abusing a dominant position. Several takeovers in the past have raised concerns regarding the interplay of consumer privacy and competition. Among those the most notable are the Google/DoubleClick Case of 2008, the FaceBook/WhatsApp Case of 2014 and most recently Microsoft's proposed purchase of LinkedIn. The

[7] There are a number of examples such as Google's Pixel Smartphone using FPC Tech, Jawbone Fitness Trackers using biometric sensors, Facebook's facial recognition technology, as well as Apple Pay using fingerprint scanners.

European Commission has now indicated that it changes policies regarding the merger of data rich-companies.[8]

Merger Control: Thresholds of Notification of Data-rich Mergers

Under the current legal framework, it is the annual turnover of the combined businesses, which needs to exceed specific thresholds in order to oblige companies to notification. In Germany the transaction volume was recently added to account for notification of mergers in cases, where turnover does not reach the notification threshold.[9] To ensure that smaller, but data-rich mergers are captured as well as envisioned by the Commission, the number of active user/customer profiles should be added, as well the differentiation power of information (identifiability).

Replicability of Datasets and Patented Technologies

In the past, is has been stated that Big Data in itself is not a problem, but once data are unique and not easily replicable, competition could be stifled.[10] Traditionally, CAs would evaluate whether two firms are direct competitors and if the merged entity commands market power post-merger. Purchasing a competitor with a similar set of data seems to *not* be the dominant mode of expansion in digital markets. Mergers here are often intended to increase the *variety* in the data portfolio a firms holds by adding *complementary data sets* or analytical tools. If data sets act as key input into downstream offerings and are not easily replicable, they may constitute a market barrier.

A merger of data-rich firms not only facilitates the merging of *two* databases, *but of any databases* the merging entities hold given that privacy policies can be changed *any time* as currently the case. Two examples illustrate this. In 2012, Google changed its Privacy Policy to merge user data from YouTube, Gmail, Google Plus, Google search and essentially also DoubleClick (Dutch Data Protection Authority 2013). The second example is FaceBook's policy change in 2016 regarding using WhatsApp data for ad

[8] Bodoni, S. (2016). EU's Vestager 'Exploring' Need to Probe Deals With Valuable Data, Bloomberg Technology, https://www.bloomberg.com/news/articles/2016-09-29/eu-s-vestager-exploring-need-to-probe-deals-with-valuable-data

[9] Reform of the Act against Restraints of Competition (Competition Act – GWB) in Germany.

[10] Kroh, E. (2016). Big Data May Warrant New EU Antitrust Rules, Vestager Says, Law360, http://www.law360.com/articles/846169/big-data-may-warrant-new-eu-antitrust-rules-vestager-says

targeting.[11] We can safely assume that the more data is linked together, the more useful it is and the more unique it becomes (Hardy 2015).

Three points can be made: (1) The *portfolio of linked databases* of these companies *cannot* be easily replicated by competitors; (2) the *sheer size in data volume, i.e. the number of users,* matters in particular when sparse datasets are compiled (de Fortuny et al. 2013); and (3) there may be also exclusivity in *access to analytical techniques* that either help to produce the data or that are used to analyze it, and these are in many cases patented technologies including algorithms.[12]

Note that in *ReedElsevier-ChoicePoint* (2008), the Federal Trade Commission argued that the combined companies provided a unique combination of quality of data (breadth and depth) and analytics that other firms would not have, which disables effective competition. In that case, however, data products were directly sold to buyers.

Dominant Position and Information Concentration

In digital competition, a dominant position in the market is gained by reaching scale in terms of data volume, scope in terms of data variety and precision in terms of analytics, apart from the other key aspects such as an innovative product or service innovation. Dominance itself is not a problem, but its misuse to avert competition. Misuse exists if a dominant company raises prices or reduces product quality or output. Such actions lower economic welfare compared to a level existing in a competitive market. But what CAs typically look for, price hikes, quality deterioration or innovation decline might be of less importance in today's digital markets.[13] Rather firms that expand will typically improve product or services quality for users and third parties like advertisers. They will also improve on micro-targeting precision, dynamic profiling of individuals and personalization of products and prices.

Same must not hold for the privacy conditions in the term of trade. In fact one of the main questions is whether deterioration in privacy conditions constitutes an abuse of

[11] Seetharaman, D. and B.R. Fitzgerald (2016). WhatsApp to Share User Data with Facebook, Wall Street Journal (Aug 25, 2016), http://www.wsj.com/articles/whatsapp-to-share-user-data-with-facebook-1472137680

[12] For example, FaceBook secures a creditworthiness-analysis method in Patent US 9100400 B2 ("Authorization and authentication based on an individual's social network"), moreover, Google holds a patent on "Systems and methods for promoting personalized search results based on personal information" (US 8620915 B1).

[13] Although such cases exist, it is rather rare that prices of freemium products are raised post-merger. Moreover, quality typically *improves* due to increased personalization of products and/or services and across the firms' portfolio. Finally, innovation is often likely to *increase* as personal data can be used across the firms' portfolio, it is not factor-specific.

market power. Such a case is currently under review at the German competition authority (Bundeskartellamt) with respect to FaceBook.

The dominant position allows the imposition of trading conditions that enable information bundling and tying, such as the aforementioned merging of different databases. While there will be efficiency gains, we cannot automatically assume that increased targeting precision always benefits consumers (Ghose and Huang 2009, Jentzsch et al. 2013). Depending on the context, personalization may reduce churn, raise switching costs and shift informational rents away from consumers. If there is no competitor holding the same *amount and variety of data* matching the data portfolio *and* analytical precision of the dominant firm, competition may in fact be stifled.

Protecting Competition by Adding Noise

Instead of relying on privacy promises by firms in lieu of a merger, CAs should condition a merger of data-rich firms on provable privacy guarantees. These guarantees would have to be specified in detail and in compliance with the Privacy Policies of the respective firms and act as commitment device. Depending on the context, different techniques and combinations thereof could be used such as randomization and/or generalization, preventing the linkability of the data, for example. These privacy guarantees could be dynamically monitored using the same mechanism as proposed above for DPAs. Such a mechanism would preventively avert any post-merger actions that are not aligned with pre-merger promises.

In Germany, there are already rules on the admissibility of the transfer of personal data in merger cases, although primarily related to the personal data of staff of the involved firm. Here, it needs to be evaluated whether any transfer of anonymized, pseudonymized or statistically aggregated data on staff ought to be transferred in due diligence procedures (Grimm 2012).

In the U.S., the FTC took action against Google, FaceBook, MySpace and others stating that these companies "deceived consumers by making commitments to limit data sharing practices and then made changes to those practices." (MIT 2013: 7). It is reported that each of these companies are now (as a result) under a 20-year consent decree with the FTC, which requires the creation of elaborate privacy procedures and subjects them to biannual audits. In applying a more nuanced and formally provable approach, it can be ensured that companies can extract some value from their data sets (such as usage patterns, etc.), without compromising individual privacy or excessively linking datasets at the individual level, creating a multi-sided platform challenge for competitors.

III Conclusions

In the era of Big Data, new tools for a credible market supervision need to be employed. Competition policies are intended to protect competition on behalf of the public. Information asymmetries are necessary for competition and data protection and privacy guarantees are critical for information asymmetries. Seen this way, they are critical for ensuring competition in digital markets.

Using privacy guarantees for supervision provides an incentive for firms to use de-personalized information to a greater extent in order to avoid scrutiny by supervisors. Moreover, such deployment could spur investments in the development of more *efficient privacy guarantees and mechanisms*.

It is argued that authorities in charge for data protection, unfair commercial practices and competition could use privacy guarantees more effectively. It is clear, though, that competition and data protection authorities would have to work together in a far more integrated way. The devil, as always, is in the detail. Considering today's firms, we are faced with differing 'publishing' scenarios, data types and dimensions, and database architectures. Moreover, there are many different metrics in development and experimentation phase. However, as more economic transactions move online, and more objects become inter-connected, we need to discuss effective approaches that align the protection of consumers and the protection of competition.

The presented approach has some key advantages: it furthers consumer privacy by limiting privacy harm, puts a spotlight on guarantees as commitment device and upholds information asymmetries needed to keep competition open.

Annex

Anonymity Anonymity of a data subject means that the subject is not identifiable or uniquely characterized within a set of subjects, the anonymity set (Diaz et al. 2002, Pfitzmann and Hansen 2010, with modifications by the author). Anonymity denotes the absence of a differentiation degree or the possibility to single out a specific individual.

> ***k-anonymity*** This concept is a privacy requirement on released data.[14] A released set of data is said to be *k*-anonymous if the information for each person contained in the released set cannot be distinguished from at least k-1 individuals who are also part of the released set (Sweeney 2002). The concept is applied to the release of datasets or sub-samples drawn from datasets.

Anonymization Anonymization is a mathematical guarantee achieved by removing or manipulating direct and indirect identifiers to irreversibly prevent re-identification (Future of Privacy Forum 2016). Different methods are applied to achieve anonymization such as k-anonymity, suppression, aggregation, perturbation or generalization. The term is sometimes used interchangeably with de-personalization or de-identification.[15]

De-personalization The term refers to "the modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labor be attributed to an identified or identifiable individual." (Federal Data Protection Act of Germany, Bundesdatenschutzgesetz). The term is sometimes used interchangeably with de-identification.

Differential privacy This concept is a formal approach that induces random noise that is large enough to hide individual contributions to the data set, but small enough to allow useful analyses (MIT 2013). It guarantees the participant of a database that the released outcome of two neighboring datasets (one with and one without that participant) are almost the same (Heffetz and Ligett 2014). "Informally the definition stipulates that any individual has a very small influence on the (distribution of the) outcome of the computation." (Talwar et al. 2013).

Identifiability Identifiability of a data subject means that the subject is identifiable within a set of subjects, meaning the data subject can be *uniquely* characterized within the subject set (Pfitzmann and Hansen 2010 with modifications). It has been stated that identifiability runs a spectrum from anonymity to full identification (Lowrance 2006). Thus, identifiability denotes a differentiation degree.

Identification Identification of a data subject means that the subject is *uniquely identified* within a set of subjects either based upon direct or indirect identifiers. ***Direct identifiers*** are attributes that specifically relate to an individual such as name, address, identification numbers, or biometric information. ***Indirect identifiers*** are attributes that in combination can identify an individual. Examples typically include date of birth, ZIP code and race, religion, financial and medical information (see also U.S. Department of Education 2016).

---

[14] The method is patented in Patent US 7269578 B2 "Systems and methods for de-identifying entries in a data source", https://www.google.com/patents/US20020169793?hl=de

[15] The word is proposed by scientist Latanya Sweeney to be used as a substitute for anonymization, as the latter is often used in a context, where effective anonymization cannot be achieved (see Ohm 2010: 1744).

Multi-party computing Multi-party computing is a cryptographic protocol that allows the computation of a function without compromising the data inputs of individual participants and without delegating the computation to a trusted third-party. The output of the function can be made public.

Pseudonymity A pseudonym is an alias that differs from a person's real name (Pfitzmann and Hansen 2010 with modifications). Pseudonyms are chosen in order to protect the real identity of a person. Pseudonymous data contains no direct identifies, but indirect identifiers remain intact (Future of Privacy Forum 2016).

Uniqueness Uniqueness relates to the differentiation power or identification quality of personal data. Given a simply anonymized mobile phone user dataset, unicity is defined as $\mathcal{E} \sim (v * h)^{-p/100}$ where spatial resolution is denoted by v (e.g. network cells) and temporal resolution by h (e.g. hours). p denotes the number of randomly chosen known spatio-temporal points (de Montjoye et al. 2013).

References

Article 29 Data Protection Working Party (2014). Opinion 05/2014 on Anonymization Techniques, adopted on 10 April 2014, Brussels, Belgium, www.cnpd.public.lu/de/publications/groupe.../wp216_en.pdf

Autorité de la concurrence and Bundeskartellamt (2016) Competition Law and Data, Gemeinsames Papier, http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2

Competition and Markets Authority (2015). The commercial use of consumer data – Report on the CMA's call for information, CMA38 June 2015.

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.

de Mulder, Y., Danezis, G., Batina, L. & Preneel, B.(2008). Identification vialocation-profiling in GSM networks (WPES'08), //researchmicrosoft.com/en-us/um/people/gdane/papers/GSMLocation-profile.pdf

de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. & Blondel, V.D. (2013). Unique in the Crowd: The privacy bounds of human mobility. Nature 3, 1376; DOI:10.1038/srep01376.

de Montjoye Y.-A., Radaelli L., Singh V. K., Pentland A. S., (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata, Science 347 (6221): 536-539.

Diaz, C., S. Stefan, J. Claessens, B. Preneel (2002). Towards measuring anonymity, PET 2002, PET'02 Proceedings of the 2nd international conference on Privacy enhancing technologies, *https://securewww.esat.kuleuven.be/cosic/publications/article-89.ps*

Dutch Data Protection Authority (2013). Investigation into the combining of personal data by Google - Report of Definitive Findings, November, z2013-00194, https://autoriteitpersoonsgegevens.nl/sites/default/files/downloads/mijn_privacy/en_rap_2013-google-privacypolicy.pdf

Future of Privacy Forum (2016). A visual guide to practical data de-identification, https://fpf.org/wp-content/uploads/2016/04/FPF_Visual-Guide-to-Practical-Data-DeID.pdf

Grimm, D. (2012). Woerz, Arbeitnehmerdatenschutz beim Betriebsübergang, ZD-Aktuell 2012, 02753, https://rsw.beck.de/cms/?toc=ZD.60&docid=327193

Hardy, S. (2015). Utility vs Privacy – Why de-identification is difficult, Presentation, Privacy and Data Protection Week, 9-13 May 2016.

Heffetz, O. and K. Ligett (2014). Privacy and Data-based Research, *Journal of Economic Perspectives* 28 (2): 75-98.

Jentzsch, N. (2016). Wettbewerbspolitik in digitalen Märkten: Sollte Datenschutz eine Rolle spielen? DIW Roundup 94 (14. April 2016), https://www.diw.de/de/diw_01.c.530874.de/presse/diw_roundup/wettbewerbspolitik_in_digitalen_maerkten_sollte_datenschutz_eine_rolle_spielen.html

Jentzsch, N. (2007). Financial Privacy – An International Comparison of Credit Reporting Systems (Springer-Verlag, Heidelberg).

Jentzsch, N., Sapi, G. and Suleymanova, I. (2013). Targeted pricing and customer data sharing among rivals, International Journal of Industrial Organization 31 (2): 131-144.

Lowrance, William W. (2006). Privacy, Confidentiality, and Identifiability in Genomic Research, Discussion document for workshop convened by the National Human Genome Research Institute, Bethesda, October 3–4, 2006.

MIT (2013). MIT Big Data Initiative at CSAIL – Workshop Report, http://bigdata.csail.mit.edu/sites/bigdata/files/u9/MITBigDataPrivacy_WKSHP_2013_finalvWEB.pdf

MIT (2014). Big Data Privacy Workshop – Advancing the State of the Art in Technology and Practice, Workshop Summary Report, http://web.mit.edu/bigdata-priv/images/MITBigDataPrivacyWorkshop2014_final05142014.pdf

Narayanan, A. und V. Shmatikov (2008). Robust De-anonymization of Large Sparse Datasets, https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, UCLA Law Review 57: 1701 – 1777.

Sweeney, L. (2002). k-Anonymity: A model for Protecting Privacy, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (5): 557 – 570, http://dataprivacylab.org/people/sweeney/kanonymity.html

Ton, A. (2014). Big Data Privacy Preservation, Ericsson Research Blog.

UK Information Commissioner (2012). Anonymisation: managing data protection risk - code of practice, https://ico.org.uk/media/1061/anonymisation-code.pdf

U.S. Department of Education (2016). Privacy Technical Assessment Center, http://ptac.ed.gov/glossary/direct-identifier