Juridisk Publikation
STOCKHOLM - UPPSALA - LUND - GÖTEBORG

EIRIK JUNGAR

Big Data
Mind the Gap – Regulation Meets Reality

Särtryck ur häfte 1/2016

Nummer 1/2016

BIG DATA

MIND THE GAP – REGULATION MEETS REALITY

By Eirik Jungar[1]

Data protection is not a new phenomenon. The current EU legislation was adopted twenty years ago. Much has happened since: the dominance of e-mail, the advent of social networks, smartphones, credit cards and everyday objects connected to the Internet. Individuals leave breadcrumbs of digital information wherever they go and whatever they do. Big data, the collection of large amounts of unstructured data that is analyzed using computer algorithms, is a way to use this information to great benefit for individuals, organizations and society. In consequence, our every move is recorded and analyzed in hope of discovering useful correlations. This raises concerns about privacy. To combat these privacy issues, the European Commission has proposed a new regulation for data protection. In this article, I examine how the existing and proposed data protection rules apply to common uses of big data and evaluate whether they can strike an adequate balance between beneficial use and privacy risks. I argue that the proposed regulation risks failing in both goals as it is built on the premise that individuals should protect their own privacy by taking control of the information they create. This may not be possible in the age of big data given the sheer amount of information generated and the unpredictability of what can be found out by analyzing the information.

1. INTRODUCTION

The Data Protection Directive[2] (DPD) revolutionized data protection by introducing binding rules on the lawful processing of personal data. That was twenty years ago, when floppy disks were the primary way of transferring digital data. Today individuals create vast amounts of data as they go about their daily lives: walking with a mobile phone, sending e-mails, using social media, making purchases, searching the web, driving through tolling booths etc.[3] Researchers have had to invent new terms to describe the volume of data

1 Law student, Uppsala University. The article is a reworked version of a paper I wrote for the course EU Commercial Law and Litigation.

2 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 *on the protection of individuals with regard to the processing of personal data and on the free movement of such data.*

3 Tene, O., Polonetsky, J. (2013) *Big Data for All: Privacy and User Control in the Age of Analytics*, Northwestern Journal of Technology and Intellectual Property, Vol 11 No 5, p 240.

generated,[4] and the amount of information is likely to grow exponentially as the Internet becomes intertwined with the physical world.[5]

The explosion of new data combined with cheap storage, efficient processing and modern data analytics heralded the advent of "big data". Big data promises big benefits for individuals, organizations and society. Amongst other things, it fuels new research, services and improved efficiency. However, there are growing concerns about privacy, our every move being recorded and analyzed. How to balance the rewards and privacy risks of big data has been called "the biggest policy challenge of our time".[6]

The European Commission (The Commission) presented a new regulation to put individuals back in control of their personal data. While some legal scholars appreciate the changes,[7] others have called it a regulatory backlash threatening innovation and beneficial use of big data.[8] The debate on data protection is held on an abstract level, discussing policy rather than regulatory models and their consequences.

The aim of this article is to analyze how the existing and proposed data protection rules apply to common uses of big data and evaluate whether they can strike an adequate balance between beneficial use and privacy risks. Can the current and proposed framework for data protection handle the challenge to privacy while reaping the promised benefits of big data?

4 Kuner, C., Cate, F.H., Millard, C. (2012) *The Challenge of 'Big Data' for Data Protection*, International Data Privacy Law, Vol 2 No 2, p 47.

5 Munir, B.A., Yasin, M.H.S., Muhammad-Sukki, F. (2015) *Big Data: Big Challenges to Privacy and Data Protection*, International Scholarly and Scientific Research and Innovation, Vol 9, p 355.

6 Tene, O., Polonetsky K. (2013) *Privacy and Big Data: Making Ends Meet*, 66 Stan.L.Rev 25, p 26.

7 See for example Spina, A. (2014) *Risk Regulation of Big Data: Has the Time Arrived for a Paradigm Shift in EU Data Protection Law?* European Journal of Risk Regulation, Volume 5 Number 2; Weibe, A. (2015) *Data Protection and the Internet: Irreconcilable Opposites? The EU Data Protection Reform Package and CJEU Case Law*, Journal of Intellectual Property Law&Practice, Vol 10 No 1; Ferretti (2014) *Data Protection and the Legitimate Interests of Data Controllers: Much Ado About Nothing or the Winter of Rights?* Common Market Law Review, Vol 51, pp 843–868.

8 See for example Ohm, P. (2013) *Response: The Underwhelming Benefits of Big Data*, 161 University of Pennsylvania Law Review Online 339; Rubinstein, L.S. (2013) *Big data: the End of Privacy or a New Beginning*, International Data Privacy Law; Tene, Polonetsky (2013) *supra* n 6; Kuner, Cate, Millard (2012) *supra* n 4; Cristensen, L., Etro, F. (2013) *Big Data, the Cloud and the EU regulation on Data Protection in EU Data Protection Reform: Opportunities and Concerns*, Intereconomics; Cate, F.H., Mayer-Schönberger, V. (2013) *Notice and Consent in a World of Big Data*, International Data Privacy Law, Vol 3 Iss 2.

First, I outline what big data is, its benefits and how it threatens privacy. The second part of this article examines how data protection rules are likely to apply to big data using three broad categories of common uses. Data protection law has until recently been a neglected area by the Court of Justice of the European Union (CJEU).[9] To examine how data protection law applies I therefore use, beyond the wording of the secondary legislation, preparatory works and opinions of the Article 29 Working Party (WP).[10] The WP is an expert organ charged with the task of enhancing a consistent application of data protection law across the union. For that purpose, the WP writes opinions on difficult legal issues, which, although not binding, are a fair indication of the correct interpretation.[11] The overarching challenge of data protection law is how to balance privacy against beneficial use. I therefore consider economic analysis to evaluate the impact of regulatory models and to some degree research in the field of statistics to understand certain privacy risks.

The conclusion is that European data protection law is flexible enough to handle the technological challenges. However, navigating the legal issues to strike a decent balance requires the agility of an acrobat. The analysis points to some fundamental problems with the current regulatory model: perhaps a braver reform is necessary to reap the benefits of big data without sacrificing our privacy.

2. BIG DATA

2.1 WHAT IS BIG DATA?

Big data refers to large amounts of different types of data, produced and stored at high speed, which is analyzed using modern technology.[12] The data analysis relies on running algorithms to discover correlations and unlike traditional research it does not require a hypothesis.[13] The findings are therefore often unintuitive and unpredictable.[14] "Google Flu Trend" is a fitting example.

9 Considering the long life of the DPD, there have been surprisingly few cases regarding its interpretation. Though recent developments show increased fervor by the court, see C-131/12 *Google Spain*; C-362/14 *Schrems*, and perhaps the less known C-230/14 *Welltimo*.

10 A working party composed of representatives from national data protection authorities. See DPD 29.

11 DPD 29, 30; also note the enhanced authority in the proposed regulation DPR 65–67.

12 Kemp, R. (2014) *Big Data and Data Protection*, White Paper, Kemp IT Law, p 2.

13 Zarsky, T.Z. (2004) *Desperately Seeking Solutions: Using Implementation Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society* 56 Maine Law Review, pp 27–28.

14 Zarsky (2004) *supra* n 13 pp 27–28.

Surprisingly, Google predicted influenza outbreaks by analyzing search-queries, which was certainly not the original intent when gathering the data.[15] Big data can change modern society, but is it for better or for worse?

2.2 THE BENEFITS OF BIG DATA

Harnessing big data can be highly beneficial for society, private organizations and individuals. The global economy stands to gain from increased productivity, innovation and efficiency.[16] Amongst other things, big data creates value by improving healthcare, traffic management, fraud detection etc.[17] For example, big data allows for research in healthcare using minimal resources.[18] Researchers were able to identify characteristics linked to "venous thromboembolic events" using data on over 900 000 patients originally gathered for billing, treatment or other research projects.[19] A study prepared by the Centre for Economics and Business Research estimates the value of big data to the UK economy being £216 billion in the next five years.[20]

Private organizations also benefit. Big data creates new business models. For example, "fintech" firms assess how a business is doing by using algorithms fed with data from social-media reviews and use of logistic firms, which allows for lending to small businesses that would be turned down by traditional banks.[21]

Big data also allows for organizations to improve their services. For example, Netflix not only collects data on what people watch, but also when they pause or switch shows, to better design their television series.[22] Google analyzes behavior to train its search algorithms, improve its translation service and to fund its operations through selling customized advertising.[23]

15 Tene, O., Polonetsky, J. (2012) *Privacy in the age of Big Data: A time for Big Decisions*, 64 Stan.L.Rev 63, p 64.
16 Cristensen, Etro (2013) *supra* n 8 p 277.
17 Munir et al (2015) *supra* n 5 p 357.
18 Schwartz, P.M. (2013) *Information Privacy in the Cloud*, University of Pennsylvania Law Review Volume 161 Number 6, pp 1631–1632.
19 Kaelber, D.C., Foster, W., Glider, J., Love, E.T., Jain, A.K. (2012) *Patient Characteristics Associated with Venous Thromboembolic Events: A Cohort Study Using Pooled Electronic Health Record Data*, 19 J. AM. MED. INFORMATICS ASS'N, p 967.
20 Center for Economics and Business Research (2012) *Data Equity: Unlocking the Value of Big Data, Executive Summary*, pp 4, 35.
21 The Economist *The Fintech Revolution*, May 9 2015.
22 Leonard, A. *How Netflix is Turning Viewers into Puppets*, Salon, Feb 1 2013.
23 The Economist *Clicking for Gold: How the Internet Companies Profit from Data on the Web*, Feb 25 2010.

For traditional business, big data can enhance efficiency and improve decision making,[24] which could increase productivity by as much as 5–6 %.[25] For example, Wal-mart's "retail link" lets suppliers keep track of their stock and sales by store and hour allowing for efficient distribution and retailing.[26] Wal-mart also discovered that consumers wanted not only flashlights and batteries after a hurricane, but also sugary snacks such as pop-tarts, an insight which allowed stores to better supply their stocks.[27] The Royal Shakespeare Company created a marketing campaign that increased regular visitors by 70 % by sifting through seven years of sales data to discover common characteristics of their best customers.[28] These improvements will trickle down to individuals, who will benefit from improved services, products and customization.[29]

To conclude, big data is beneficial and will perhaps reshape society in ways one cannot fully predict. However, a sober approach is warranted when evaluating the benefits: what betterments are worth having to, perhaps, give up some of our privacy to realize?[30]

2.3 THE PRIVACY RISKS OF BIG DATA

Though highly beneficial, big data also threatens privacy. Understanding the privacy risks is vital for evaluating the lawfulness of the processing of personal data since privacy is one of the primary objects of data protection law.[31] While the Charter of Fundamental Rights of the European Union (the Charter) recognizes protection of personal data as a separate right,[32] its role is still understood in terms of privacy protection.[33] When the CJEU rules on the

24 McKinsey Global Institute (2012) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, p 5.

25 Rubinstein (2013) *supra* n 8 p 3.

26 Tene, Polonetsky (2012) *supra* n 15 p 64.

27 The Economist *A Different Game: Information is Transforming Traditional Businesses*, Feb 25 2010.

28 The Economist *A Different Game: Information is Transforming Traditional Businesses*, Feb 25 2010.

29 Tene, Polonetsky (2013) *supra* n 6 p 28.

30 Ohm (2013) *supra* n 8 pp 339–346.

31 DPD recital 2, 10.

32 Charter art 8.

33 See for examples on efforts to conceptualize the right to personal data protection De Hert, P., Gutwirth, S. (2006) *Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power* in Claes, E., Duff, A., Gutwirth, S. (2006) *Privacy and the Criminal Law*, Intersentia; Poullet, Y., Rouvroy, A. (2009) *The Right to Informational Self-Determination and the Value of Self-Development: Reassesing the Importance of Privacy for Democracy* in Gutwirth, S., Poullet, Y., De Hert, P., Terwangne, C. D., Nouwt, S. (2009) *Reinventing Data Protection?* Springer; Tzanou, M. (2013) *Data Protection as a Fundamental Right Next to Privacy? 'Reconstructing' a not so New Right,* International Data Privacy Law, Vol 3, No 2; Lynksey, O. (2014) *Deconstructing*

lawfulness of processing of personal data it cites both the right to private life and protection of personal data, stressing the right to respect for private life.[34] I therefore firstly outline the concept of privacy and secondly the privacy harms of big data. Privacy has been understood as first the right to seclusion or opacity, second in terms of non-interference in decisions belonging to the individual and third in terms of control of personal information.[35] Privacy can therefore be conceptualized as an instrument to protect the autonomy and development of one's person free from undue interference by ensuring a private space and providing tools for individuals to live and present themselves as they see fit.[36]

The privacy risks of big data can be compartmentalized into two categories: harms relating to collection and harms relating to the use of data. Both were noted in the case *Digital Rights Ireland* concerning the illegality of the Data Retention Directive[37] due to the vast amount of data collected and stored.[38]

Firstly, increasingly obscure ways of collecting data online can in itself be harmful. Not knowing when information is gathered can result in a feeling of being under constant surveillance.[39] Individuals may therefore be forced to live under the presumption that their every move is observed and recorded.[40] This hinders personal development and autonomy as people act differently depending on who is watching.[41] There is a risk that individuals will avoid engaging in certain behavior altogether.[42]

Secondly, there are privacy risks relating to the use of data. The more data is gathered, the more revealing each piece of data becomes.[43] Combinations of harmless data can be used to infer sensitive information about an individual.[44]

Data Protection: The Added-value of a Right to Data Protection in the EU Legal Order, International and Comparative Law Quarterly 63.

34 See e.g. from before the Charter was binding C-465/00, C-138/01 and C-139/01 *Rundfunk* para 68; C-275/06 *Promusicae* para 63–65 and after C-92/09 and C-93/09 *Volker* para 47, 52; C-131/12 *Google Spain* para 81, 99; C-362/14 *Schrems* para 39.

35 Poullet, Rouvroy (2009) *supra* n 33 pp 61–62.

36 Poullet, Rouvroy (2009) *supra* n 33 pp 75–76.

37 Directive 2006/24/EC of the European Parliament and the Council of 15 March 2006 *on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.*

38 C-293/12 and 594/12 *Digital Rights Ireland.*

39 C-293/12 and 594/12 *Digital Rights Ireland* para 37.

40 Spina (2014) *supra* n 7 p 251.

41 Lynksey (2014) *supra* n 33 pp 589–590.

42 Tene, Polonetsky (2013) *supra* n 3 p 256.

43 C-293/12 and 594/12 *Digital Rights Ireland* para 27; Tene, Polonetsky (2013) *supra* n 3 p 251.

44 Lynksey (2014) *supra* n 33 p 252.

For example, Target Inc. accurately predicted a woman's due date based on her purchasing habits, amongst other things a preference for un-scented lotion and tendency to buy zinc and magnesium supplements.[45] This harms privacy since individuals may not have wished the inferred information to be disclosed in such a context and therefore inhibit the free development of their persona.[46] For example, a person belonging to a sexual minority may not wish their sexual orientation to affect certain social relations. Yet, there is a risk that the person's sexuality will be inferred and disseminated in an unwanted context as it has been shown to be quite easy to deduce a person's sexuality by analyzing their behavior on social networks.[47] Predictions based on correlations can be harmful in more concrete ways, for example to discriminate.[48] What if an employer or insurance company could infer that a particular job candidate or consumer is more likely to develop cancer or Alzheimer's disease based on data acquired from a third party?

2.4 A QUEST FOR BALANCE – REFORM EFFORTS OF THE EU

Considering the potential rewards and privacy risks of big data, most scholars, stakeholders, and policymakers agree that there is a need for a regulation that allows for beneficial use of big data while minimizing the impact on privacy.[49] The debate concerns how to strike the right balance. The current framework for data protection, the DPD, is showing signs of age. The Commission has proposed to replace the DPD with a new regulation on data protection (DPR).[50] The Commission observes that the sheer volume of data combined with an unawareness of how information is collected and used result in a growing concern about privacy online.[51] On June 24 2015 the European Parliament, Council and Commission entered a trilogue to reach a mutually accepted text.[52] An agreement was struck on December 15 allowing for a formal adoption

45 Tene, Polonetsky (2013) *supra* n 3 p 253.

46 Nissenbaum, H. (2004) *Privacy as Contextual Integrity*, 79 Washington Law Review 119, p 155.

47 Cristensen, Etro (2013) *supra* n 8 p 277.

48 Tene, Polonetsky (2013) *supra* n 3 p 253.

49 See e.g. Tene, Polonetsky (2012) *supra* n 6 p 67; Cristensen, Etro (2013) *supra* n 8 p 280; Cate, Mayer-Shönberger (2013) *supra* n 8 p 71; Ohm (2013) *supra* n 8 pp 340–341.

50 COM (2012) 11 Final, Proposal for a Regulation of the European Parliament and of the Council *on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*.

51 COM (2012) 9 Final, Communication from the commission to the European Parliament, the Council, the European economic and social committee and the committee of the regions *Safeguarding Privacy in a Connected World: A European Data Protection Framework for the 21st Century* p 4; Special Eurobarometer 359 (2011) *Attitudes on Data Protection and Electronic Identity in the European Union.*

52 OJ C301/1 *EDPS recommendations on the EU's options for data protection reform.*

early 2016.[53] The final text is yet to be released, but the compromise draft of the regulation resulting from the trilogue is available.[54] The compromise text does not change the DPR substantially in relation to the subject dealt with in this article, but I will discuss the differences between the compromise text and the DPR where it is relevant.

The Commission holds the objectives and principles of the current framework to still be relevant, but technological developments demand a stronger protection of personal data.[55] The main thrust of the proposed regulation, as far as big data is concerned, is to provide tools for individuals to take control of their personal data by narrowing the scope for valid consent,[56] clarifying the principles of data minimization and purpose limitation,[57] and enhancing transparency through notification requirements and simplified access.[58] The response to more data is simply more of the same kind of protection.[59] Some appreciate the increased rights of data subjects,[60] others call it a regulatory backlash[61] that risks dampening the possible rewards of big data.[62] The debate is often abstract, omitting an analysis of how the rules may actually apply to common uses of big data. In the following section, I examine the application of the DPD and DPR to uncover where they are lacking.

3. DATA PROTECTION AND BIG DATA

3.1 COMMON USES OF BIG DATA

The uses of big data are diverse, consisting of basically all analytics of large quantities of data. Certain kinds of use have the same legal issues in common. I therefore broadly categorize three common uses: non-personal,

53 IP/15/6321 European Commission, Press Release 15.12.2015 *Agreement on Commission's EU Data protection reform will boost Digital Single Market.*

54 Interinstitutional File 2012/0011 (COD) *Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation, Analysis of the final compromise text with a view to agreement)*.

55 COM (2012) 11 *supra* n 50 p 2; COM (2012) 9 supra n 51 p 3.

56 DPR recital p 25; compare DPR 6(1)a and DPR 7 to DPD 7(1) a.

57 DPR recital p 30, 40; compare DPR article 5 to DPD article 6.

58 DPR recital p 38, 55; compare DPR 5(1) a to DPD 6(1) a and DPR 15 to DPD 12.

59 Gilbert, F. (2012) *EU Data Protection Overhaul: New Draft Regulation*, The Computer and Internet Lawyer, Vol 29 No 3, p 2.

60 E.g. Weibe (2015) *supra* n 7, pp 65–66; Lynksey (2014) *supra* n 33 pp 594–595; WP *Opinion 01/2012 on the Data Protection Reform Proposals* pp 2–6.

61 Tene, Polonetsky (2012) *supra* n 15 p 63.

62 Rubinstein (2013) *supra* n 8 pp 2–7; Tene, Polonetsky (2012) *supra* n 15 p 67; Cate, Mayer-Schönberger (2013) *supra* n 8 p 67; Cristensen, Etro (2013) *supra* n 8 pp 277–280; Spina (2014) *supra* n 7 p 252; Munir et al (2015) *supra* n 5 pp 359–360.

semi-personal, and personal uses of big data. The non-personal category concerns data without connection to individuals, such as predicting the weather from temperature readings.[63] This category falls outside the scope of this article since data protection rules do not apply to non-personal data.

The semi-personal category concerns data derived from the behavior of individuals used to discover knowledge of general value. Drawing from the examples above, this includes Wal-mart's retail link and hurricane snacks, Netflix monitoring to improve its television series, health research and Google's use of data to train translation and search engine algorithms. What semi-personal uses have in common is that the utility of the data does not necessitate individuals to be identified. Therefore, anonymization has been a key strategy for lawful processing.[64] The central legal issues are *when* and *how* data protection law applies to anonymized datasets.

The personal category concerns data derived from individuals, analyzed to discover knowledge of a specific person. Targeted behavioral advertising is the most important use in this category. Examples range from retailers offering coupons based on previous purchases to the Royal Shakespeare Company's marketing campaign to even more opaque ways of collecting data on the internet using cookies and other forms of tracking to profile users. In these situations, data protection law will always apply. The key legal issue is if there is an appropriate legal ground for processing of personal data.[65]

3.2 THE SEMI-PERSONAL CATEGORY

3.2.1 SCOPE OF APPLICATION AND PITFALLS OF ANONYMIZATION

Both the DPD and the DPR apply to the processing of *personal data*, that is, data that can be linked to an identifiable individual.[66] If information that can identify an individual is removed, generalized or randomized it falls outside the scope of data protection law and organizations can use and trade the data freely. Anonymization techniques have been a key strategy to reap the rewards of big data.[67] However, recent studies show that anonymized datasets may not

63 See for example *IBM's recent purchase of The Weather Company, Mourdoukoutas, IBM to buy The Weather Company*, Forbes Oct 28 2015.

64 WP *Opinion 5/2014 on Anonymization Techniques* p 5.

65 The topical issue of big data use for national security purposes also falls within this category, though it will not be analyzed in this paper due to the fundamentally different legitimizing purposes for processing.

66 DPD 3, 2 (1) a; DPR 2, 4 (1) (2).

67 WP *Opinion 5/2014 supra* n 64 p 5.

be as anonymous as previously believed.[68] *When* and *how* data protection law applies to anonymized data is a tricky legal question and important to get right for the overarching question of balance. For organizations, the possibility of anonymized data falling within the scope of the DPD and DPR risks outlawing beneficial uses and imposing big compliance costs. For individuals, the security of anonymization is important for privacy protection. It is impossible to control personal information, utilizing such tools as access, deletion, and right to object, when neither data controller nor subject know whose data is processed.

Data is always personal if it contains direct identifiers such as a name or photo.[69] Data protection rules also apply when different pieces of information can be combined to distinguish an individual.[70] For example, the IP-address does not directly identify an individual, but it is unique to a computer and the machine can easily be tied to its owner (the exception being, for example, public computers in libraries).[71] The DPR and DPD apply to data that can be tied to a person indirectly if a person can be identified by *means likely reasonably to be used by the controller or any other individual*.[72] There must be more than a negligible possibility of identification.[73] What if Google sells a packet of search queries where personal identifiers are removed and replaced by a code? What if health researchers create a dataset with information on disease, symptoms, gender, date of birth and where the patient was treated? Do they have to comply with data protection rules?

Three years ago, the answer would probably have been no. However, recent research challenges the security of anonymization. The search engine AOL released queries from 650 000 users where personal identifiers had been removed and replaced by a code. Some were singled out and identified by the uniqueness of their searches. A 62-year old widow was tracked down and became known as the woman who had searched for "numb fingers", "60 single men" and "dog that urinates on everything".[74] Latanya Sweeney, a statistician, discovered that general information in combination often uniquely identify an

68 Ohm, P. (2010) *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA Law Review, p 1701 ff.
69 DPD 2a; DPR 4 (1).
70 WP *Opinion 4/2007 on the Concept of Personal Data* p 13; DPD 2a; DPR 4(1).
71 WP *Opinion 4/2007 supra* n 70 p 14 see also C-191/01 *Bodil Lindqvist* para 27 *"…identifying them by name or by other means, for instance by giving their telephone number or information regarding their working conditions and hobbies"*.
72 DPD recital p 26; DPR recital p 24.
73 WP *Opinion 4/2007 supra* n 70 p 15.
74 Ohm (2010) *supra* n 68 pp 1717–1718 (She had "made the searches on behalf of friends").

individual. Using a 5-digit ZIP-code, gender and date of birth 87 % of the US population could be identified; 53 % were likely to be uniquely identified by only place, gender and date of birth.[75] In the "Netflix Prize Data Study", researchers could identify the individuals behind anonymized film-reviews by linking the data released by Netflix to data available publicly online on the review-site IMDb.[76] Analyzing common anonymization techniques, the WP concluded that *all common anonymization techniques are flawed in that they can allow for re-identification*.[77]

There is always a risk of re-identification. But if the anonymization is secure enough it may not be possible with means *reasonably likely* to be used. How secure must anonymization be for it not to contain personal data? Several factors support a broad interpretation. The WP states that re-identification must be reasonably impossible for it not to be considered as personal data.[78] The intent of the European legislator was that it should be construed "as general as possible".[79] The data protection rules intend to protect fundamental rights, and should therefore not be interpreted restrictively.[80] Even scholars who believe the risk of re-identification to be exaggerated contend that it will always be possible to identify *some* individuals, but with robust anonymization it will likely stay in the 1–3 percentile.[81] The wording of DPD and DPR suggests that for applicability it does not matter that only a few can be identified. Data is personal if it can identify *an individual*.[82] If some individuals are easily distinguished from a dataset, the whole set contains personal data.

What means are *reasonably likely* to be used by the data controller or any other person? The first question regards *whom* the test looks at as different

75 Sweeney, L. (2000) *Simple Demographics Often Identify People Uniquely*, Carnegie Mellon University, Data Privacy Working Paper 3, p 2.

76 Narayanan, A., Shmatikov, V. (2008) *Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)*, The University of Texas at Austin, pp 1, 8–14.

77 WP *Opinion 5/2014 supra* n 64 pp 23–24.

78 WP *Opinion 5/2014 supra* n 64 pp 6, 8.

79 WP *Opinion 4/2007 supra* n 70 p 4; see the commentaries on art 2 in COM (1992) 422 Final *amended proposal for a COUNCIL DIRECTIVE on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, p 10 and OJ C93/1 *COMMON POSITION (EC) No 1/95 adopted by the Council on 20 February 1995 with a view to adopting Directive 95/…/EC of the European Parliament and of the Council of … on the protection of individuals with regard to the processing of personal data and on the free movement of such data* p 20.

80 C-131/12 *Google Spain* para 53.

81 Golle, P. (2006) *Revisiting the Uniqueness of Simple Demographics in the US Population,* Palo Alto Research Center, p 1; Cavoukian, A., Castro, D (2014) *Big Data and Innovation, Setting the Record Straight: De-identification Does Work*, Information and Privacy Commissioner, p 4.

82 DPD 2a, recital 26; DPR 4(1), recitals 23–24.

individuals and organizations have different means available. The WP interprets *any other person* literally. If, for example, a data controller creates an anonymized dataset that it passes on but keeps an identifiable version, the anonymized dataset is personal data in the hands of all ensuing controllers since the original dataset can easily identify individuals.[83] Given such an interpretation, almost all data is personal since there is always someone with specific information or tools that easily identify an individual.[84] Such an interpretation risks arbitrary results. A data controller who has no knowledge of the original collector and relies on robust anonymization to process data would break the law without a possibility of knowing it. I suggest that *any other person* should be interpreted as a standardized test. Data would be personal if identification is possible considering data easily obtained, publicly available or already in possession of the controller.

Re-identification becomes easier the more data is made available.[85] A substantial amount of anonymized data will be considered personal and fall within the scope of DPD and DPR. This raises the question of *how* the rules apply. Across the Atlantic, the increased scope of the DPD is cited as a warning example of all-encompassing and onerous requirements threatening use and innovation, even by pro-privacy scholars.[86] Is this necessarily the case: can it not permit beneficial uses while protecting privacy?

3.2.2 CONSEQUENCES OF APPLICABILITY

For the processing of personal data to be lawful the data controller must rely on a *legal ground,*[87] and comply with the *principles of lawful processing*.[88] The "balance of interest provision" is the most important legal ground, since one cannot ask for consent of the data subject when he or she is not identified. Under the balance of interest provision, processing is legal provided that it strikes the right balance between the legitimate interests of the controller and the fundamental rights and freedoms of the data subject.[89] The privacy risks are small and processing should be lawful if robust anonymization limits the potential of re-identification to a small number of individuals. However, the assessment may be difficult to make in practice. First, there are no guidelines

83 WP *Opinion 5/2014 supra* n 64 p 9.
84 Roosendaal, A. (2014) *Digital Personae and Profiles in Law: Protecting Individuals' Rights in Online Contexts*, Wolf Legal Publishers, pp 179–180.
85 Ohm (2010) *supra* n 68 pp 1756–1757.
86 Ohm (2010) *supra* n 68 pp 1762–1763.
87 DPD 7; DPR 6.
88 DPD 6; DPR 5.
89 DPD 7f; DPR 6(1)f.

on which anonymization techniques are considered secure.[90] It is therefore difficult to estimate the privacy risks. Second, evaluating the benefits is hard. The result of the data analysis is unpredictable; it may discover correlations that improve a service or help cure diseases but it is as likely to be utterly useless.[91]

A possible solution is for the Commission to adopt acceptable standards of anonymization that are specific to certain industries and common uses, which is possible under the proposed regulation.[92] The standards could take into account the expected benefits of common uses (such as market research or health research) and the character of the data (sensitive or non-sensitive) and set an anonymization standard accordingly. If a dataset is anonymized according to the standard, organizations could rely on the balance of interest provision to process data lawfully. This allows for the balancing of utility against privacy risks without imposing undue costs for organizations in making complex assessments.

However, the balance of interest provision cannot be relied on in all circumstances. Sensitive data, for example concerning an individual's health, may only be processed with consent of the data subject.[93] It is impossible to obtain consent if one does not know who the data subject is. Also, re-identification is an act of processing in itself and would therefore require consent to be lawful.[94] If the DPD and DPR applies to anonymized data and processing requires consent then there are no possibilities of lawful processing. Such a total prohibition fails to strike an adequate balance between utility and privacy, especially considered the broad interpretation of sensitive data.[95] The DPR provides a possible solution and way around the prohibition in an exemption from the consent requirement for necessary processing for *historical, statistical or scientific research.*[96] A similar exemption from the purpose limitation principle existed in the DPD.[97] The WP held that *statistical purposes* covers a wide range of activities, including analytical tools of websites and big data

90 WP *Opinion 5/2014 supra* n 64 pp 23–24.
91 Roosendaal (2014) *supra* n 84 p 189.
92 Made possible in the proposed DPR 23(4) and 87(2).
93 DPD 8; DPR 9.
94 DPD 2b; DPR 4(3).
95 See C-101/01 *Bodil Lindqvist* para 51, information about an injured foot constituted sensitive data concerning health.
96 DPR 9(2)i.
97 DPD 6(1)b.

applications aimed at market research.[98] The new provision could prove to be a solution to the consent conundrum as there is no indication that a change of scope was intended by adding *research*.

3.2.3 THE PURPOSE LIMITATION PRINCIPLE

The balance of interest provision can provide a legal ground for processing anonymized data. For processing to be legal, data controllers also need to comply with principles of lawful processing.[99] The principle of purpose limitation states that personal data can only be collected for specified, explicit, and legitimate purposes, and may not be used in a way incompatible with the purpose for which it was collected.[100] The purpose must be specific, which means that general purposes such as "improving user experience" or "future research" are not adequate.[101]

It may be difficult for semi-personal uses of big data to comply with the principle. First, the findings of the data analytics are difficult to predict.[102] Being more specific than "improving efficiency" or "developing services" may not be possible. Second, excluding data originally gathered for other purposes risks dampening innovation. For example, the medical research on venous thromboembolic events relied on data originally gathered for billing purposes and other projects.

The purpose limitation principle is built on the premise of control of one's personal data; when data is used in a manner not foreseen by individuals when disclosing information, it violates their expectations of privacy.[103] Is this still the case when the individual is not identified? Should not actual harm to privacy be the relevant limitation on use, not the original purpose? An exemption from the principle may be warranted for anonymized data.

The DPD provides an exemption from the purpose limitation principle for *processing of data for historical, statistical or scientific purposes*.[104] The WP argues that *statistical purposes* covers analytical tools of websites and big data applications aimed at market research.[105] However, in the DPR the exemption is

98 WP *Opinion 03/2013 on Purpose Limitation* p 29, pp 45–47.
99 DPD 6; DPR 5.
100 DPD 6(1)b; DPR 5 b.
101 WP *Opinion 03/2013 supra* n 98 p 15–16.
102 Zarsky (2004) *supra* n 13 pp 27–28.
103 WP *Opinion 03/2013 supra* n 98 p 4.
104 DPD 6b.
105 WP *Opinion 03/2013 supra* n 98 pp 29, 45–47.

removed and replaced by a sub-paragraph to the legal ground provision stating "where the purpose of further processing is not compatible with the one for which data have been collected, the processing must have a legal basis at least in one of the grounds referred to in points (a) to (e) of paragraph 1".[106] In essence, this is an exception to the purpose limitation principle when relying on any legal ground except the balance of interest provision. The paragraph has been criticized by the WP,[107] the European Data Protection Supervisor,[108] and the European Parliament,[109] because a new legal ground is necessary irrespectively since the requirements of a legal ground and compliance with the principles are cumulative.[110] Regardless of whether the regulation is passed as it stands or with the provision removed, processing based on the balance of interest provision that is incompatible with the original purpose will always be unlawful.[111] This would seriously curtail beneficial use of big data even when it is subject to robust anonymization. It is also contradictory to introducing an exemption from the consent requirement for sensitive data processed for statistical, historical or scientific research purposes as such processing will often be unlawful if there is no similar exception from the purpose limitation principle.

The compromise draft of the DPR,[112] resulting from the trilogue, addresses some of the issues of the purpose limitation principle. The compromise text adopts an exemption to the purpose limitation principle similar to the one in the DPD. It is provided in the definition of the principle that processing for, amongst other things, statistical research, is not considered incompatible with the original purpose.[113] This solution is preferable to the one originally proposed and analyzed above. Firstly, it retains the logical consistency of the European data protection framework in that the requirement of a legal ground and compliance with the principles are cumulative. Secondly, it solves the inconsistency pointed out above where processing for scientific purposes was exempt from the requirement of consent but not the purpose limitation

106 DPR 6(4).
107 WP *Opinion 03/2013 supra* n 98 pp 36–37.
108 Opinion of the European Data Protection Supervisor, *on the Data Protection Reform Package* 07.03.2012, p 20 para 121–124.
109 COM (2012) 0011C7-0025/2012 – 2012/0011(COD) *European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).* The EP proposed to remove the provision.
110 DPD 6; DPR 5; WP *Opinion 03/2013 supra* n 98 pp 36–37.
111 As the provision is found in DPR 6(1)f and therefore not excluded by 6(4).
112 Interinstitutional File 2012/0011 (COD) *supra* n 54.
113 CDPR 5(1)b, recital 40.

principle. Thirdly, the exemption may allow for adequate processing of semi-personal big data if *statistical research* is interpreted broadly to include marketing and commercial purposes. However, is returning to the solution in the DPD satisfactory for today's data protection? The original purpose is less fit to differentiate between legitimate and illegitimate processing when the scope of data protection extends to most anonymized data and consequentially a different purpose may not, in itself, be harmful to privacy.

3.2.4 IMPLICATIONS FOR PRIVACY PROTECTION AND ANCILLARY REQUIREMENTS

The problem of consent is symptomatic of a more fundamental issue: the data protection rules were not designed to apply to data when data controllers must not know whom the data concerns. Consent, notice, right to access and deletion are the cornerstones of data protection,[114] but none of these can be exercised without knowing whom to ask for consent or whom to inform. The proposed regulation solves these issues for organizations processing data by the introduction of a new provision. Article 10 of the DPR states that if the processed data does not permit the controller to identify a natural person, they are not required to do so for the sole reason of complying with any provision of the regulation.[115] The other side of the story, how to protect privacy when an individual cannot take control of their personal data, is not addressed.

3.3 THE PERSONAL CATEGORY

3.3.1 THE PROMISES AND PERILS OF TARGETED ADVERTISING

The personal category concerns big data used to uncover information about a specific individual and will always fall within the scope of the EU data protection law.[116] I focus on targeted/behavioral advertising, that is, collecting data to create a profile of an individual which is used to sell customized advertisements.[117] An appropriate regulation of online targeted advertisement is important for the purpose of balancing utility and privacy risks of big data. It fuels the provision of new services by allowing personal data to be used as a source of revenue, creating a market for personal data,[118] but there are privacy risks.

114 See e.g. DPD recital 25; DPR recital 6, 7, 9.
115 DPR 10, see also recital p 45.
116 Relevant provisions in the E-privacy Directive, 2002/58/EC should be noted, which under 5(3) requires consent for tracking by cookies.
117 Berger, D.D. (2011) *Balancing Consumer Privacy with Behavioral Targeting*, Santa Clara Computer & High Technology Legal Journal 3, pp 6–7.
118 Tene, O., Polonetsky, J. (2012) *To Track or "Do Not Track": Advancing transparency and Individual Control in Online Behavioral Advertising*, Minn.J.L.Sci.&Tech. 281, p 341.

Collection of personal data on the internet is often done in obscure ways more or less unknown to the data subject and used for opaque purposes.[119] For example, to create an accurate profile, Google will record users' searches, which links they click, their e-mail activity on Gmail and which videos they watch on YouTube.[120] Facebook uncovers similar information from embedded "like" buttons on websites, what users share, which events they attend etc.[121] It is difficult to keep track of how personal data is used since data brokers in the business of collecting, analyzing and selling personal information are increasingly common.[122] And the more data that is gathered about an individual, the greater the risk of sensitive inferences.[123] The most recent Eurobarometer on the topic concludes that very few (15 %) respondents felt in control of their data and that most people (67 %) were concerned about this.[124]

3.3.2 STRENGTHENING THE REQUIREMENT OF CONSENT

Because of the privacy implications, online targeted advertisement has been in the crosshairs of the WP and EU. Due to the difficulty for individuals to understand how their data is collected and used, the WP holds that processing personal data for behavioral online advertisement normally requires consent and should not be lawful under the balance of interest provision.[125] For the same reasons, the DPR purports to put individuals back into control by requiring consent to be explicit and forcing data controllers to provide more information on when, how and for what reasons data is processed.[126] Similarly, some scholars have criticized a liberal approach to the balance of interest provision, arguing that it creates a loophole in the protection of personal data.[127]

119 Castelluccia, C. (2012) *Behavioral Tracking on the Internet: A Technical Perspective* Chapter 2 in Gutwirth, S., Leenes, R., De Hert, P., Poullet, Y. (2012) *European Data Protection: In Good Health?*, Springer, p 23–27; Richards, N.M., King, J.H. (2013) *Three Paradoxes of Big Data*, 66 Stan.L.Rev. Online 41, p 42; Chester, J. (2012) *Cookie Wars: How New Data Profiling and Targeting Techniques Threaten Citizens and Consumers in the "Big Data" Era* Chapter 4 in Gutwirth, S., Leenes, R., De Hert, P., Poullet, Y. (2012) *European Data Protection: In Good Health?*, Springer, pp 53, 59–60.

120 Castelluccia (2012) *supra* n 119 pp 23–27.

121 Roosendal, A. (2012) *We Are All Connected to Facebook… by Facebook!* Chapter 1 in Gutwirth, S., Leenes, R., De Hert, P., Poullet, Y. (2012) *European Data Protection: In Good Health?*, Springer, pp 4 –7.

122 Roosendaal (2014) *supra* n 84 pp 180–181.

123 See above, section 2.2.

124 Special Eurobarometer 431 (2015) *Data Protection Report*, p 6.

125 WP *Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller Under Article 7 of Directive 95/46/EC p 68; WP Opinion 03/2013 supra* n 98 p 45.

126 Compare DPD 12a and DPR 15; see COM (2012) 11 *supra* n 50 p 2; COM (2012) 9 *supra* n 51 p 3.

127 Ferretti (2014) *supra* n 7 pp 858, 867–868.

The argument behind requiring and strengthening consent is that individuals should be given a real choice of participation as a lack of control may result in an unwillingness use new technologies, services and share information to the detriment of consumers and business alike.[128] This reasoning is questionable on two counts. First, providing individuals with more information does not necessarily mean they make more informed decisions. Empirical studies show that users do not read, nor understand privacy notices.[129] This is understandable: it would take approximately 30 working days per year to read all the privacy notices we encounter.[130] Simply providing a feeling of security and control results in individuals disclosing more information irrespective of whether it correlates to actual security or control.[131] The challenge to consent is a practical one: individuals may neither have the time nor the will to take control of their personal data, no matter which tools are provided. Second, this means that consent for processing of personal data for online targeted advertising, even if given, is unlikely to be valid.

Consent is a ground for lawful processing expressly recognized in the Charter.[132] Data protection is closely related to privacy and if individuals exercise their autonomy by consenting to their data being used, there is less intrusion in their privacy.[133] However, this is only the case if the individual understands what consent entails. In a context where individuals do not understand the consequences of consenting, accepting consent as a legal ground could in practice *weaken* the position of the data subject by legitimizing such processing.[134] The DPD and DPR address this issue in the definition of consent. To be valid, consent must be *freely given, specific, informed* and, in the case of DPR, *explicit.*[135]

128 Roeber, B., Rehse, O., Knorrek, R., Thomse, B. (2015) *Personal Data: How Context Shapes Consumers' Data Sharing with Organizations from Various Sectors*, Electronic Markets 25, p 105; Spiekermann, S., Acquisti, A., Böhme, R. Hui K. (2015) *The Challenges of Personal Data Markets and Privacy*, Electronic Markets 25, p 16.

129 Cate, H.F. (2006) *The Failure of Fair Information Practice Principles, Chapter 13 in Winn, J.K. Consumer Protection in the Age of Information Economy*, Ashgate, pp 361–363; Eurobarometer 431 *supra* n 124 pp 84–86.

130 McDonald, M. A., Cranor, L.F. (2008) *The Cost of Reading Privacy Policies*, A Journal of Law and Policy for the Information Society, p 17.

131 Brandimarte, L., Acquisti, A., Loewenstein, G. (2013) *Misplaced Confidences: Privacy and the Control Paradox*, Social Psychological and Personality Science, Vol 4, Iss 3, pp 345–346.

132 DPD 7a; DPR 6 a; Charter art 8(2).

133 Roosendaal (2014) *supra* n 84 pp 15–16.

134 WP *Opinion 15/2011 On the Definition of Consent*, p 10.

135 DPD 2 h; DPR 4(8).

To be *freely given*, there must be an absence of deception, coercion or significant negative consequences if a data subject does not consent.[136] If processing of personal data is not necessary for the provision of a service, requiring an individual to disclose information to receive the service will in general invalidate consent.[137] In the compromise text resulting from the trilogue, it is expressly made clear that consent is not, in most circumstances, freely given when consenting to processing is a requisite to the provision of a service.[138] Free consent may therefore be impossible for online targeted advertising. Many collectors rely on personal data as a primary source of revenue; sharing information is therefore often a requirement for utilizing the service.[139] Individuals will therefore accept to disclose data even if they would not have liked to.[140] Perhaps that is why there is a very weak correlation between privacy concerns and participation in social networks.[141]

Specific consent means that it must refer to a well-defined and concrete scope of processing, as such a blanket consent for "use for commercial purposes" is not valid.[142] Consent is valid only for purposes reasonably foreseen by the data subject when consenting.[143] When data is used to infer surprising and unintuitive information, such as Target Inc's pregnancy prediction, it will likely fail the specificity test.

Informed consent requires the data subject to understand the implications and consequences of consent and controllers must provide accurate and full information in an understandable way.[144] Obtaining informed consent is tricky: the findings of data analytics are sometimes surprising and it is difficult to consent to, for example, "analytics for marketing purposes", when one does not know what may be discovered.[145] Furthermore, it is doubtful whether individuals

136 WP *Opinion 15/2011 supra* n 134 pp 12, 34.
137 Roosendaal (2014) *supra* n 84 p 186; WP *Opinion 15/2011 supra* n 134 p 12.
138 CDPR 7(4), recital 34.
139 Chang, A., Kannan, P.K., Whinston, B. (1999) *The Economics of Freebies in Exchange for Consumer Information on the Internet: An Exploratory Study*, International Journal of Electronic Commerce, Vol 4 No 1, pp 85–87; Tene, Polonetsky (2012) *supra* n 118 p 335.
140 Tene, Polonetsky (2012) *supra* n 118 p 333.
141 Acquisti, A., Gross, R. (2006) *Imagined Communities: Awareness, Information Sharing and Privacy on the Facebook*, pp 36–58 in Golle, P., Danezis, G 2006 *Privacy enhancing technologies, 6th International Workshop, PET 2006, Cambridge, UK, June 28–30, 2006, Revised Selected Papers*, pp 56–57.
142 WP *Opinion 15/2011 supra* n 134 p 17.
143 C-543/09 *Deutsche Telekom AG* para 65, referring to valid consent within the E-Privacy directive.
144 WP *Opinion 15/2011 supra* n 134 pp 19–20.
145 Rubinstein (2013) *supra* n 8 p 5.

benefit and make more informed decisions about their privacy simply by being provided with more information and more consent boxes to tick when surfing the web.[146] In general, individuals defer to the already set "default" solution.[147] The change in DPR to require *explicit*, opt-in consent will therefore have considerable consequences. Individuals that have not opted out of disclosing personal data will not opt in either, even if it is in their best interest (unless it is required for the provision of a free service,[148] in which case the consent is unlikely to be valid anyway).

In conclusion, the scope of consent is narrow and processing big data for online targeted advertising will in general fail the test of free, informed and specific consent. Yet, the WP holds that consent should be required and the proposed reforms to strengthen data protection focus on consent as a key to put individuals back in control of their personal data. It is doubtful if such a stance improves privacy protection. Making informed decisions about one's privacy is near impossible due to the inherent characteristics of big data: the large amount generated and the unpredictable results of data analysis. Studies show that individuals do not take control of their data, and providing better tools for doing so is unlikely to change that. If consent is not a viable nor desirable legal ground to process data, is the balance of interest provision a better alternative?

3.3.3 THE BALANCE OF INTEREST PROVISION: A BETTER ALTERNATIVE?

It is often impossible to obtain valid consent for the use of big data. This means that most targeted advertising must rely on the balance of interest provision. This requires analyzing the legitimate interest (where for example fundamental rights and community goods are considered more important than commercial interests), the impact on the data subject (if the data is sensitive, if processing involves large amounts or profiling and the reasonable expectations of the data subject) and what safeguards the controller has implemented such as anonymization, transparency, and opt-out mechanisms.[149]

Traditional use of targeted advertising, such as loyalty cards, are generally acceptable if data subjects are provided with an easy way to opt out.[150] Concerning online targeted advertising, the WP is more restrictive, in principle

146 Cate (2006) *supra* n 129 pp 361–363; Eurobarometer 431 *supra* n 124 pp 84–86.
147 Acquisti, A., John, L.K., Loewenstein, G. (2013) *What Is Privacy Worth?*, The Journal of Legal Studies, Vol 42 No 2, pp 268–269.
148 Tene, Polonetsky (2012) *supra* n 118 p 333.
149 WP *Opinion 06/2014 supra* n 125 pp 50–51.
150 WP *Opinion 06/2014 supra* n 125 p 59.

demanding consent.[151] However, obtaining *valid* consent is practically not possible. Given the WP's interpretation of the balance of interest provision, online targeted advertising would in practice be prohibited due to there being no possible legal ground. Such a conclusion is not necessary. The balance of interest provision should be a possible legal ground since online targeted advertising brings important benefits for consumers and corporations alike.

Concerning the legitimate interest, one should consider the wider interests involved in online targeted advertising and not strictly a controller's interests in advertisement revenue. Corporations who purchase advertisements also have legitimate interests. Industry stakeholders note that small companies, which cannot afford wide marketing campaigns, would not be commercially viable if not for affordable and effective advertising.[152] Many organizations rely on personal data as a source of revenue to create content and services, other organizations utilize the data to advertise, improve and provide new products.[153] The market in personal data can bring big benefits: the Boston Consulting Group puts the annual economic benefit of digital personal profiles at 330€ billion by 2020.[154] Data subjects also stand to gain, firstly from the fact that content, such as search engines, mobile applications and social networks, are provided for free (to be more precise, for the price of disclosing personal data).[155] Secondly, targeted advertising reduces search costs for acquiring goods and services.[156]

There are privacy risks. The obscurity of data collection online means that it is difficult to know when one is under surveillance, which could harm individual autonomy and have a chilling effect on behavior.[157] However, the solution should not be to demand uninformed consent. Rather, these issues should be addressed by demanding transparency from data processors about how they collect and process data.[158] If individuals are aware of when their behavior is recorded the privacy implications are less severe since individuals are not

151 WP *Opinion 06/2014 supra* n 125 p 68; WP *Opinion 03/2013 supra* n 98 p 45.
152 Federation of European Direct and Interactive Marketing, Data industry platform (2011) *Proposal for a Balanced Approach on Consent,* pp 7–8; Zarsky (2004) *supra* n 13 pp 33–34.
153 Berger (2011) *supra* n 117 pp 30–32.
154 Boston Consulting Group (2012) *The Value of our Digital Identity*, Liberty Global Policy Series, p 21.
155 Though it is difficult for individuals to determine the value of their data and consequently if they are making a good deal in disclosing it in exchange for the service, see Spiekermann, Acquisti, Böhme, Hui (2015) *supra* n 128 p 163.
156 Chang, Kannan, Whinston (1999) *supra* n 139 p 95.
157 See above, section 2.2.
158 Tene, Polonetsky (2013) *supra* n 3 p 270–271.

forced to presume constant surveillance. Also, subject to public scrutiny, organizations may avoid unethical and intrusive processing.[159] Harms pertaining to use regarding big data concern the possible sensitive inferences from compiling information from different sources.[160] And while it is true that there is a bigger potential of misuse the more data is available, that is not a reason to outlaw processing for acceptable uses. Admittedly, telling acceptable and unacceptable apart is not an easy task,[161] but the guideline must be actual privacy harm, not the risk of misuse.

The value created by online advertisement fuels the provision of new kinds of services based on a trade in personal data to a modest detriment to privacy.[162] The proposed interpretation of the WP, to require opt-in consent, risks outlawing the budding industry based on online targeted advertisement and trade in personal data because consent for such purposes is unlikely to be valid even if given. There are real and serious privacy risks. These can be managed within the balance of interest provision by requiring a data controller to establish proper safeguards. Data subjects should have the possibility to know when their information is being collected and they should be given an opportunity to object, which means that organizations engaged in targeted advertising must be transparent about the ways in which they collect and analyze personal data and provide simple ways to opt out.

4. CONCLUDING REMARKS

The aim of this article is to examine how European data protection law applies to common and important uses of big data to evaluate whether it can strike the right balance between beneficial use and privacy risks. To analyze the application, I distinguish between three uses of big data. The non-personal uses which fall outside the scope of the DPD, DRP and this article. The semi-personal, when big data derived from individuals is utilized to discover knowledge of general value and the personal where big data is used to uncover information about a specific person.

In the semi-personal category, the problematic legal issue is how the data protection framework applies to anonymized datasets. The more data is made available, the easier re-identification becomes. The scope of European data protection rules will expand and encompass most anonymized datasets, which risks outlawing beneficial use and imposing compliance costs on organizations.

159 Tene, Polonetsky (2013) *supra* n 3 p 270–271.
160 See above, section 2.2.
161 Cate, Mayer-Schönberger (2013) *supra* n 8 p 69.
162 Tene, Polonetsky (2012) *supra* n 118 p 341.

Some of the issues of this development can be alleviated by the Commission adopting implementing acts under the proposed regulation to set up acceptable standards of anonymization that strike the right balance of privacy risks against big data rewards. However, such efforts will be futile if the purpose limitation principle is interpreted in a strict way and the exemption for statistical purposes is removed.

In the personal category, online targeted advertising comes to the foreground as the petrol that fuels many beneficial uses of big data by allowing personal data to be used as a source of revenue. Due to the use of obscure collection methods and risk of misuse of profiles, the EU, WP and some scholars believe that processing of personal data for the purpose of such advertising should require consent. Such an interpretation risks outlawing online targeted advertisement to the detriment of organizations and consumers alike as it is difficult to obtain valid consent.

In conclusion, the data protection framework is flexible enough to handle the challenge of big data. This requires sensible interpretations of primarily the balance of interest provision and purpose limitation principle taking into account the wider interests at stake. The analysis of the legal issues points to a more fundamental problem at the core of data protection law. The DPD and DPR are, to a large degree, built on the premise that individuals should control their own personal information and protect their own privacy. However, in the age of big data it is increasingly difficult to take control of one's personal data and empirical studies show that individuals are not likely to make informed decisions about their privacy. When data protection laws apply to anonymized data, none of these controls can be exercised. The main thrust of the reform, which focuses on strengthening individual control, is therefore unlikely to improve privacy protection while restricting the utility of big data.

The Commission holds that the principles and objectives of the DPD are still as relevant as they were twenty years ago and that big data is no different from other data. The conclusion of this article puts a question mark after that statement. While an umbrella may keep you dry in a drizzle, people prefer to stay inside when it pours. Merely strengthening the principles and objectives of the DPD is akin to giving individuals a bigger umbrella, and expecting them to gladly face a storm. There is a risk that individuals prefer to stay inside rather than face the flood of information or read the never-ending privacy policies to

control the data they create, which would rob them and businesses many of the benefits of modern life. Protecting privacy in the age of big data requires a rethinking of data protection. Another kind of data protection is necessary, but it falls outside the scope of this article to investigate more radical alternatives. Hopefully an understanding of the merits and shortcomings of the current regulation will help in the evaluation of other options.