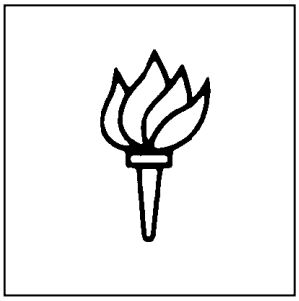
NEW YORK UNIVERSITY SCHOOL OF LAW

PUBLIC LAW & LEGAL THEORY RESEARCH PAPER SERIES
WORKING PAPER NO. 12-56



Big Data: The End of Privacy or a New Beginning?

Ira S. Rubinstein

October 2012

Big Data: The End of Privacy or a New Beginning?

Ira S. Rubinstein*

'Big Data' refers to novel ways in which organizations, including government and businesses, combine diverse digital datasets and then use statistics and other data mining techniques to extract from them both hidden information and surprising correlations. While Big Data promises significant economic and social benefits, it also raises serious privacy concerns. In particular, Big Data challenges the Fair Information Practices (FIPs), which form the basis of all modern privacy law. Probably the most influential privacy law in the world today is the European Union Data Protection Directive 95/46 EC (DPD).[1] In January 2012, the European Commission (EC) released a proposal to reform and replace the DPD by adopting a new Regulation.[2] In what follows, I argue that this Regulation, in seeking to remedy some longstanding deficiencies with the DPD as well as more recent issues associated with targeting, profiling, and consumer mistrust, relies too heavily on the discredited informed choice model, and therefore fails to fully engage with the impending Big Data tsunami.

My contention is that when this advancing wave arrives, it will so overwhelm the core privacy principles of informed choice and data minimization on which the DPD rests that reform efforts will not be enough. Rather, an adequate response must combine legal reform with the encouragement of new business models premised on consumer empowerment and supported by a personal data ecosystem. This new business model is important for two reasons: First, existing business models have proven time and again that privacy regulation is no match for them. Businesses inevitably collect and use more and more personal data, and while consumers realize many benefits in exchange, there is little doubt that businesses, not consumers, control the market in personal data with their own interests in mind. Second, a new business model, which I describe below, promises to stand processing of personal data on its head by shifting control over both the collection and use of data from firms to individuals. This new business model arguably stands a chance of making the FIPs efficacious by giving individuals the capacity to benefit from Big Data and hence the motivation to learn about and control how their data are collected and used. It could also enable businesses to profit from a new breed of services

Key Points

- Big Data—which may be understood as a more powerful form of data mining that relies on huge volumes of data, faster computers, and new analytic techniques to discover hidden and surprising correlations—challenges international privacy laws in several ways: it casts doubt on the distinction between personal and non-personal data, clashes with data minimization, and undermines informed choice.
- Europe is presently considering a General Data Protection Regulation that would replace the ageing Data Protection Directive. This Regulation both creates new individual rights and imposes new accountability measures on organizations that collect or process data.
- But the Big Data tsunami is likely to overwhelm these reform efforts. Thus, a supplementary approach should be considered using codes of conduct. In particular, regulators should encourage businesses to adopt new business models premised on consumer empowerment by offering incentives such as regulatory flexibility and reduced penalties.

* Ira S. Rubinstein is Senior Fellow and Adjunct Professor of Law, Information Law Institute, New York University School of Law. Email: ira.rubinstein@nyu.edu. I am grateful to Microsoft Corporation for supporting the research for this paper; however, the views expressed herein are solely those of the author.

1 Directive (EC) 95/46 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 [1995] OJ L281/31.

2 '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)' COM(2012) 11 final ('Regulation').

that are both data-intensive and imbued with privacy values.

This paper has three parts. The first part examines Big Data in greater detail and asks whether it defeats traditional privacy law by undermining core principles and regulatory assumptions. If so, regulators need to consider new approaches beyond those reflected in their current thinking. The second part analyses whether the proposed Regulation successfully addresses the challenges of Big Data; this includes a close examination of the new and revised provision on automated decision making or profiling. The thrid part begins by describing a 'control shift' that is already underway due to the emergence of a new business model based on 'Personal Data Services' or PDSes. Next, it considers to what extent PDSes are technologically feasible, intellectually coherent, and realistic from a business standpoint. Finally, it offers a few preliminary recommendations on how EU institutions might foster PDSes by granting regulators more authority to experiment with codes of conduct. In particular, regulators should consider offering incentives to firms that adopt this new business model ranging from regulatory flexibility to reduced penalties.

The EU Directive and the Big Data challenge

Core privacy principles

The DPD sets out core privacy principles relating to 'personal data', that is, information about an identified or identifiable person. These principles have the goal of permitting only legitimate processing of personal data. They include principles of data quality (characterized in terms of purpose limitation, data minimization, accuracy, and completeness),[3] consent,[4] transparency,[5] access and rectification,[6] confidentiality,[7] and security.[8] Beyond these core principles, the DPD also seeks to ensure the free flow of personal data within the EU and addresses transfer of personal data to third countries, jurisdictional rules, administrative matters, and enforcement.

There are three main shortcomings with the DPD's core privacy principles. First, as Professor Fred Cate has argued, the DPD relies too heavily on informed choice.[9] This is problematic given that empirical studies show individuals neither read nor understand privacy policies, which anyway rely on ambiguous language, and are easily modified by firms. Thus, consent is too often an empty exercise. Second, while the DPD also requires data minimization, there are relatively few instances in which data protection authorities have forced technology firms to re-design their software, hardware, or business processes to minimize the processing of data or make it possible for data subjects to use such systems anonymously. Third, the DPD has failed to keep pace with globalization, the relentless improvement and expansion of technological capabilities, and the changing ways in which individuals create, share, and use personal data.[10] To state the obvious, the DPD is showing its age, having preceded the commercialized Internet, the World Wide Web, laptops, mobile computing, GPS, RFID, and Web 2.0 services, not to mention Big Data.[11]

Recent EU reform efforts

In 2010, the EC published a Communication concluding, inter alia, that while the core principles of the DPD were still valid, the Directive could no longer meet the challenges of 'rapid technological developments and globalisation'.[12] The Communication briefly mentions far reaching changes such as the popularity of social networking sites that permit individuals to voluntarily share personal data, the growth of cloud computing, the ubiquity of mobile devices and of physical sensors that transmit geo-location information, and the growing use of data mining technologies enabling the aggregation and analysis of data from multiple sources. Nevertheless, the Commission's initial response to these changes emphasized standard data protection measures such as enhancing consent, strengthening transparency, and clarifying and making more explicit certain preconditions of data protection including data minimization and the right of access.

The Commission then engaged in a consultation process that culminated in the publication of a Regulation in January 2012. It is clear that in developing a new set of data protection rules, the Commission was well aware of the 'dramatic technological changes'

3 Article 6.

4 Articles 2(h) and 7(a).

5 Articles 10 and 11.

6 Article 12.

7 Article 11.

8 Article 17.

9 Fred H Cate, 'The Failure of Fair Information Practice Principles,' in Jane K Winn (ed.), *Consumer Protection in the Age of the Information Economy* 360 (Burlington: Ashgate, 2006).

10 N Robinson et al., 'Review of the European Data Protection Directive' (2009) RAND Technical Reports 12–19 (TR–710), <http://www.rand.org/pubs/technical_reports/TR710.html> accessed 17 December 2012.

11 Christopher Kuner et al., 'The Challenge of "Big Data" for Data Protection' (2012) 2 International Data Privacy Law 47, 48.

12 European Commission Communication, 'A comprehensive approach on personal data protection in the European Union' COM (2010) 609 final.

that have occurred since the DPD was first proposed, and very concerned with problems raised by profiling and data mining.[13] Despite this realization, the Commission held firm in its belief that 'the current framework remains sound as far as its objectives and principles are concerned'.[14] While the Regulation introduces several new privacy rights—notably the right to be forgotten and to data portability—the other changes it makes are incremental at best. For example, it proposes stricter transparency obligations and a tighter definition of consent. It strengthens the existing provision concerning 'automated individual decisions' by including a new provision on profiling, but the changes are limited and focus mainly on enhancing transparency. It also imposes new responsibilities on data controllers including data protection by design and default. All of these changes are briefly discussed below.[15] While many of these new measures are well considered, and help bolster the DPD's core privacy principles, I argue below that the Regulation continues to rely on informed choice as the primary tool for addressing the new issues implicit in Big Data.

Big Data's challenge to data protection

'Big Data' (BD) is best understood as a more powerful version of knowledge discovery in databases or data mining, which has been defined as 'the nontrivial extraction of implicit, previously unknown, and potentially useful information from data'.[16] Data mining enables firms to discover or infer previously unknown facts and patterns in a database. It relies not on causation but on correlations that arise from the application of non-public algorithms to large collections of data. Consequently, the newly discovered information is not only unintuitive and unpredictable, but also results from a fairly opaque process.[17] Indeed, BD may be thought of as data mining on steroids.

The McKinsey Global Institute (MGI) recently defined BD as 'datasets whose size is beyond the ability of typical database software to capture, store, manage, and analyze'.[18] All of the biggest Internet companies—Google, Facebook, Amazon, eBay, Microsoft, and Yahoo!—are engaged in Big Data in one form or another and treat data as a major asset and source of value creation. Google is an especially good example as it relies on the availability of the data it collects from its own services not only to fund its operations (by determining and delivering relevant search ads) but also to train its search algorithms and develop new data-intensive services such as voice recognition, translation, and location-based services.[19] But BD encompasses a much wider swath of enterprises than these Internet giants, and now extends to any company (or government agency) that relies on statistical methods and data mining algorithms to analyse large datasets and thereby improve decision making, enhance efficiency, and, according to a recent study, increase productivity by as much as 5–6 per cent.[20] Indeed, there is evidence that BD has led to major breakthroughs in healthcare, more efficient delivery of electrical power, reductions in traffic congestion, and vast improvements in supply chain management.[21] BD also directly benefits consumers by enabling innovative, data-driven services such as Amazon 'Customers Who Bought This Item Also Bought' and Microsoft Fare Tracker. More generally, the MGI report finds that BD can create 'significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating substantial economic surplus for consumers'.[22] Indeed, case studies in the MGI report posit that BD will generate $300 billion of value per year in the US healthcare industry, €250 billion of value per year in European public sector administration, $100+ billion of additional revenue for service providers of location data, 60 per cent increases in net margins across the retail industry, and up to a 50 per cent decrease in product development and assembly

13 Impact Assessment (including annexes) accompanying the proposed Regulation and the proposed Directive, SEC (2012) 72 final, 24–25.

14 Ibid, at 7.

15 Other important changes, which are beyond the scope of this paper, include expanding the territorial scope of EU data protection law; addressing cross-border data transfers; introducing a new regime of penalties and fines; and shifting power from local data protection authorities to Brussels. For an overview of the Regulation, see Viviane Reding, 'The European Data Protection Framework for the Twenty-First Century' (2012) 2 International Data Privacy Law 119.

16 U M Fayyad et al., 'From Data Mining to Knowledge Discovery, An Overview,' in U M Fayyad et al. (eds), *Advances in Knowledge Discovery and Data Mining* 6 (Menlo Park: AAAI, 1996), cited in Tal Z Zarsky, 'Mine Your Own Business! Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion' (2003) 5 Yale Journal of Law and Technology 1.

17 Tal Z Zarsky, 'Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society' (2004) 56 Maine Law Review 13.

18 McKinsey Global Institute, 'Big Data: The Next Frontier for Innovation, Competition, and Productivity' 1 (May 2011).

19 John Markoff (26 June 2012) 'How Many Computers to Identify a Cat? 16,000' NY Times B1.

20 Erik Brynjolfsson *et al.*, 'Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?' (April 2011), <http://ssrn.com/abstract=1819486> accessed 17 December 2012.

21 Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics,' (forthcoming) Northwestern Journal of Technology and Intellectual Property.

22 McKinsey Global Institute (n 18), at 1–2.

costs in manufacturing. In these recessionary times, figures of this magnitude can hardly be ignored.

BD has three defining features.[23] The first is the availability of data at a massive scale collected not only online but through the use of mobile devices with location tracking capabilities and thousands of 'apps' that share data with multiple parties, interactions with smart environments (sometimes referred to as ambient intelligence or the internet of things),[24] monitoring systems in the physical environment,[25] and the human body itself, which is being used not only to harness data for genetic testing but also for authentication via biometric data.[26] Additionally, Web 2.0 services enable users to create and voluntarily share vast amounts of personal data about themselves and their friends and family. Although individuals mostly volunteer these data for social purposes, organizations are happy to collect and profit from their analysis.[27] The second defining feature is the use of high speed, high-transfer rate computers, coupled with petabytes (ie millions of gigabytes) of storage capacity, resulting in cheap and efficient data processing. This increasingly means reliance on the cloud-computing model.[28] The third and final feature is the use of new computational frameworks (such as Apache Hadoop) for storing and analysing this huge volume of data.

In light of these three features, it is impossible to overstate the vast profusion of digital data now available to organizations or the novel ways in which BD combines these diverse datasets. Nor is it surprising that BD should intensify existing privacy concerns over tracking and profiling. With the advent of BD, cookies, and Web beacons are no longer the primary culprits. Rather, profiling technologies now extend to every aspect and phase of individual and social life, with BD supplying the necessary horsepower to find hidden correlations and make interesting predictions, some of which may benefit individuals or society, while others may be more problematic.

Policy considerations

BD raises a number of policy considerations, with privacy scholars tending to highlight two main issues: privacy and discrimination. Due to space constraints, this paper limits itself exclusively to privacy, leaving the equally important issues regarding discrimination for another day.[29]

Privacy

BD challenges the very foundations of the DPD (and all similar privacy laws) by enabling re-identification of data subjects using non-personal data, which weakens anonymization as an effective strategy, thereby casting doubt on the fundamental distinction between personal data and non-personal data.[30] BD also greatly exacerbates the dignitary harms associated with amassing information about a person—what Professor Daniel Solove refers to as aggregation.[31] With its massive scale, continuous monitoring from multiple sources, and sophisticated analytic capabilities, BD makes aggregation more granular, more revealing, and more invasive. Of course, re-identification only heightens the harms associated with aggregation by enabling data controllers to link even more information to an individual's profile, leading to what Ohm calls the 'database of ruin'.[32] BD also raises a related issue concerning automated decision making, which relegates decisions about an individual's life—such as credit ratings, job prospects, and eligibility for insurance coverage or welfare benefits—to automated processes based on algorithms and artificial intelligence.[33] Not surprisingly, BD intensifies the use of automated decision making by substantially improving its accuracy and scope. Because decisions based on data mining are largely

23 Paul Ohm, 'Big Data and Privacy' (2011) unpublished paper (attributing the power of Big Data to 'more data, faster computers, and new analytic techniques'); also see Kuner (n 11), at 47.

24 See Mireille Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era' in J Bus et al. (eds), *Digital Enlightenment Yearbook 2012* 45–46 (Amsterdam: IOS Press, 2012), <http://works.bepress.com/mireille_hildebrandt/40> accessed 17 December 2012.

25 'A Special Report on Managing Information: Data, Data Everywhere' (27 February 2010) The Economist 12–13, <http://www.economist.com/node/15557443> accessed 17 December 2012.

26 Omer Tene, 'Privacy: The New Generation' (2011) 1 International Data Privacy Law 15, 19–21.

27 Ibid, at 22–25.

28 Tene and Polonetsky (n 21).

29 Data mining has been associated with three distinct forms of discrimination: Price discrimination, manipulation of threats to autonomy, and covert discrimination. There is a large literature on this topic including a number of important contributions from the early days of data mining. For an overview, see Solon Barocas, 'Data Mining: An Annotated Bibliography' (2011) Cyber-Surveillance in Everyday Life: An International Workshop, <http://www.digitallymediatedsurveillance.ca/wp-content/uploads/2011/04/Barocas_Data_Mining_Annotated_Bibliography.pdf> accessed 17 December 2012.

30 Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701. For a critique of Ohm's position and a proposal to replace the PII/non-PII distinction with a tripartite, risk-based distinction, see Paul M Schwartz and Daniel J Solove, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 N.Y.U. Law Review 1814, 1879–83.

31 Daniel J Solove, 'A Taxonomy of Privacy' (2006) 154 Penn. Law Review 477, 506.

32 Ohm (n 30).

33 Tene and Polonetsky (n 21), at 12.

invisible to their subjects, significant issues arise around access to, and the accuracy and reliability of, the underlying data. Article 20 of the Regulation is highly relevant in this regard, and we analyse it more closely below.

The impact of Big Data on data protection law

In addition to the privacy issues highlighted above, BD has an even broader impact on data protection laws such as the DPD. Recall that the DPD fundamentally relies on transparency and consent to ensure that users make informed choices about sharing personal data with organizations. While the DPD includes other substantive obligations (purpose and use restrictions, data quality, security, and access) other than security, they have limited impact because they depend on an individual's awareness that her data are being processed. Data mining and BD worsen this problem by exploding the core premises of informed choice in three ways. First, firms that rely on data mining may find it impossible to provide adequate notice for the simple reason that they do not (and cannot) know in advance what they may discover. Second, it follows that since users lack knowledge of potential correlations, they cannot knowingly consent to the use of their data for data mining or Big Data analytics. Third, privacy laws apply solely to personal data, that is, to data relating to an identified or identifiable person. But it is not at all clear whether the core privacy principles of the Regulation—transparency, consent, data minimization, access, as well as the new rights to be forgotten and to data portability—apply to newly discovered knowledge *derived* from personal data, especially when that data has been anonymized or generalized by being transformed into group profiles, that is, profiles that apply to individuals as members of a reference group, even though a given individual may not actually exhibit the property in question.[34]

BD also calls into question three longstanding regulatory assumptions of privacy laws, including the DPD. The first is whether the personal data/non-personal data distinction remains viable. As just noted, data mining extracts new knowledge from personal and non-personal data, thus creating a regulatory dilemma: should the DPD cover not only personal data but also any non-personal data that forms the basis for data extractions of new knowledge and that (once created) would be regulated as personal data? If so, there are potentially no limits to the scope of the DPD; if not, data mining may largely escape regulatory oversight, even though it permits inferences of previously private information and/or the use of group profiles that may cause as much or more harm as the regulated collection and use of personal data.[35] The second is whether anonymization—the process of removing identifiers to create anonymized data sets—remains effective in protecting users against tracking and profiling. Over the past few years, there have been several notorious cases of re-identification of individuals by cross-referencing anonymized data sets with a related set of data that includes identifiers. As already noted, BD heightens the problem by drawing on more data, faster computers, and improved analytic techniques.[36] The third is whether data minimization—the idea that personal data processing must be restricted to the minimum amount necessary—can survive the onslaught of Big Data. Simply stated, data minimization is inimical to the underlying thrust of BD, which discovers new correlations by applying sophisticated analytic techniques to massive data collection, and seeks to do so free of any *ex ante* restrictions. Because data minimization requirements would cripple Big Data and its associated economic and social benefits, regulators should expect to see this requirement largely observed in the breach.[37]

Does the Regulation succeed in addressing these Big Data challenges?

The Regulation introduces changes designed to enhance individual control over personal data by strengthening the transparency and consent provisions, revising the profiling provision, and announcing new individual rights. In addition, it enlarges the responsibilities of controllers and processors through a host of new accountability provisions.[38] Do these reforms address the challenges posed by Big Data?

Enhanced control over personal data and new responsibilities for controllers

Strengthening transparency and consent

Articles 11 and 14 propose stricter transparency obligations, requiring that information addressed to data

34 Anton Vedder, 'KDD: The Challenge to Individualism' (1999) 1 Ethics & Information Technology 275, 277.

35 For example, the credit or healthcare risks of people living in a certain neighbourhood may be higher than those in other neighbourhoods, which may result in a denial of credit or health insurance coverage for these individuals, even though a specific person living in this neighbourhood pays her bills on time and has a clean bill of health. This is not a new observation; see Vedder (n 34).

36 See Ohm (n 30) (citing AOL's release of search data and the Netflix prize dataset).

37 See Tene and Polonetsky (n 21), at 20.

38 For a similar grouping, see Reding (n 15), at 124–27.

subjects should be 'easily accessible' and 'easy to understand' and listing in great detail the types of information that controllers must provide when collecting personal data. Articles 4(8) and 7(1) propose tighter definitions of consent by clarifying that it must be not only freely given, specific, and informed, but 'explicit'—thus neither silence nor inactivity can constitute valid consent. Moreover, controllers bear the burden of proving that data subjects consented to the processing of their personal data. Even though these changes may be well taken, it is hard to imagine that they will overcome the longstanding deficiencies of the informed choice model or bring about a new era in which consumers understand their rights and act on them. My contention is simple if radical: the informed choice model is broken beyond any regulatory repair, and the only way to reinvigorate it is by changing the relevant information markets.

Profiling

Article 20 of the Regulation replaces Article 15 of the DPD, the provisions most directly relevant to profiling.[39] But there are shortcomings in the original version, and they mostly remain in the new one. Article 15(1) of the DPD addresses 'automated individual decisions' and grants the right to every person 'not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.'[40] Article 15(2) provides exceptions where a decision is taken in the course of entering into or performing a contract, and certain conditions are met or 'authorized by a law which also lays down measures to safeguard the data subject's legitimate interests'. This provision does not prohibit the creation of profiles but only how they are applied. Article 15 seems to have been motivated by the twin concerns that automated decision making will diminish the role of persons in influencing decisions affecting them and that such decisions are given too much deference (as if the mere fact that they result from sophisticated computer processes makes them more objective).[41] The three main problems with Article 15 are its limited scope and that it grants only a limited right and a limited remedy.[42]

As to scope, Article 15(1) applies only if all of the relevant conditions are met. This makes it inapplicable whenever a person exercises some level of influence over a decision-making process. Ambiguities in the meanings of several key terms (eg, 'decision', 'significantly', 'solely', 'certain personal aspects') may further limit its scope. This includes the derogations in Article 15(2), which allow automated decisions to be made about a person if 'suitable measures' are made (by contract or by law) to safeguard her 'legitimate interests'. The limited right Article 15 grants is to resist automated decisions and seek human intervention. But this is not a right to nullify such decisions (for example, unless the data subject consents to them). Rather, this right to object only comes into play if a data subject knows that he or she is subject to such decisions and lodges an objection. As we have seen from the preceding analysis, privacy harms associated with Big Data occur without the data subject's knowledge or awareness, leading to the criticism that the invisibility of profiling makes Article 15 a 'paper dragon'.[43] Nor does Article 12(a)'s right to discover the 'knowledge of the logic' of automated processes cure this problem since it, too, only benefits someone who is aware of being profiled. Finally, as to remedy, if a data subject is profiled or subjected to discrimination, Article 15 provides limited comfort. At most, it requires that the data controller bring some human judgement to bear on a decision by reviewing the factors forming the basis for the automated decision. As Bygrave notes: 'The controller is neither required to change these criteria or factors, nor to supplement them with other criteria/factors.'[44]

Article 20 of the Regulation modifies Article 15 in several ways. For example, it characterizes automated decision making more broadly, offers a new exception based on consent, prohibits automated processing based solely on sensitive data, and empowers the Commission to adopt delegated acts to further specify the nature of 'suitable measures to safeguard the data

39 They have no counterpart in US privacy law.

40 Additionally, Art. 12(a) grants the right to every data subject 'to obtain from the controller knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in Article 15(1).' This right is not absolute, however. Recital 41 suggests the relevant limitations, which depend on balancing respect for 'trade secrets or intellectual property and in particular the copyright protecting the software' against 'the data subject being refused all information'.

41 European Commission, 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, COM (92) 422 final-SYN 287, 26 (1992).

42 Lee A Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) 17 Computer Law & Security Report 17, 18.

43 Mireille Hildebrandt, 'Who is Profiling Who? Invisible Visibility,' in S Gutwirth et al. (eds), *Reinventing Data Protection?* 248 (Amsterdam: Springer, 2009).

44 Bygrave (n 42), at 20.

subject's legitimate interests'. Apart from these changes, however, its major thrust is very much the same as the earlier version.

Although Hildebrandt is critical of Article 20, she also suggests that if the notice obligation in Article 20(4) 'does not hinge on a person requesting it, this would make a radical difference with the present level of protection'. Additionally, she describes the obligation to provide notice of 'envisaged effects' as a 'revolutionary novel legal requirement'.[45] This seems overstated. First, as Hildebrandt concedes, the wording of Article 20(4) is ambiguous and may support the opposing interpretation (that notice is only due if requested). Second, whether or not notice hinges on a person requesting it, the controller's obligations under Article 20(4) apply only 'to cases referred to in paragraph 2', that is, to derogations involving entering into or performing certain contracts, statutory measures, or consent. Thus, the notice obligations are far from universal. Finally, although Hildebrandt interprets Article 20 as 'forcing data controllers to notify us what risk we are taking by leaking our data', it is hard to see why individuals are any more likely to read or understand notices about the existence and envisaged effect of profiling than they have been about other privacy notices, especially given the novelty, complexity, and obscurity of profiling to the average Internet user.

New individual rights

Article 17, the right to be forgotten and to erasure, is a highly controversial provision that builds on the existing right to deletion of data (Article 12 of the DPD) and seeks to address more effectively the privacy and dignitary harms (including reputational damage) associated with the dissemination and hence persistence of voluntarily shared data in social networking and other Web 2.0 services. Article 17(2) would require controllers to take 'reasonable steps, including technical measures', to inform third parties when a data subject has requested the erasure of previously published personal data relating to them, but this could prove burdensome or even impossible in any number of scenarios. Moreover, the right to be forgotten is not only somewhat vague and impractical as drafted but raises serious and possibly irresolvable conflicts with rights of free expression.[46] For present purposes, the key point is that the right to be forgotten is limited by its terms to personal data. Thus, it is not even clear whether Article 17 would apply to predictive inferences based on personal data that may have been anonymized or generalized as a result of analytic techniques at the heart of Big Data.

Article 18, which creates a new right to data portability, serves the highly laudable goal of enabling individuals to extract their personal data (personal profiles, photos, postings, contact lists, etc) from one application or service and move them to another as long as the data are 'processed by electronic means and in a structured and commonly used format'. While it may be undesirable for the Commission to specify these formats or any related technical standards as authorized by Article 18(3),[47] ensuring data portability remains a critical step. Data portability is a key factor in promoting competition among existing services and preventing 'lock-in'. At the same time, it greatly eases the way for the creation of new services such as the personal data services described below.

New responsibilities of data controllers

Article 22 summarizes the responsibilities of controllers including general obligations such as documentation (Article 28), data security (Article 30), impact assessments (Article 33), prior authorization or consultation (Article 34), and designation of a data protection officer (Article 35). In addition, Article 23(2) creates a more specific obligation for controllers to implement mechanisms to ensure, by default, that data minimization requirements are satisfied. (This new requirement of data protection 'by design and default' is very promising but much depends on how it is implemented.) All of these new provisions are intended to ensure that controllers process data in compliance with the core privacy principles of the Regulation. But if these core provisions fail to address Big Data satisfactorily, it is not at all clear that these additional obligations will remedy this shortcoming.

In sum, the Big Data trend—more data, faster computers, and new analytic techniques—poses severe challenges to data protection law, which has not only failed to keep pace with technological change but is even more likely to fall behind when confronted with Big Data. Even the Regulation, despite laudable efforts to

45 Hildebrandt (n 43), at 51 (citing language in Article 20(4) requiring notice of 'the existence of processing' for a measure based on profiling 'and the envisaged effects of such processing on the data subject').

46 See Center for Democracy and Technology (CDT), 'Analysis of the Proposed Data Protection Regulation', 28 March 2012, available at <https://www.cdt.org/files/pdfs/CDT-DPR-analysis.pdf> accessed 17 December 2012. In its analysis, CDT gives the example of a blogger commenting on a political controversy who might have to delete her post if it incorporates a statement by a public figure who later regrets what he said and requests its removal.

47 Ibid.

shore up certain shortcomings of the DPD, fails to alter this verdict. What then are the alternatives?

Some new ideas for addressing Big Data

Several of the authors referenced above offer new ideas for addressing the privacy implications of BD.[48] For our purposes, the most relevant idea is that of Tene and Polonetsky, who evaluate and largely reject a host of traditional legal responses (including consent, data minimization, and access). Instead, they propose a 'sharing the wealth' strategy premised on data controllers providing individuals with access to their data in a usable format and allowing them 'to take advantage of applications to analyze their own data and draw useful conclusions' from it.[49] They argue that this 'featurization' of data will unleash innovation and create new business opportunities. Their proposal is important for two reasons. First, they insist that in view of the serious privacy challenges raised by Big Data and the difficulties in regulating it, organizations should be prepared to share with individuals the wealth their data helps create. As they note, both fairness and efficiency rationales support such access and use rights for consumers.[50] Second, they recognize that 'access in a usable format' creates value to individuals and is therefore very likely to re-engage consumers who have 'remained largely oblivious to their rights'. Justice Louis Brandeis famously observed, 'sunlight is the best disinfectant'. But Tene and Polonetsky rightly point out that sunlight is not always enough, especially when 'individuals do not care for, and cannot afford to indulge in transparency and access for their own sake'.[51] Rather, transparency and access only become salient when consumers have the ability to use and benefit from their own personal data in a tangible way.[52]

All of the above recommendations are important and deserving of further consideration. In the final section of this essay, I limit myself to laying out a more fully developed version of Tene and Polonetsky's 'sharing the wealth' strategy. They suggest that a fairness rationale justifies the 'featurization' of Big Data 'regardless of whether or not you accept a property approach to personal information'. In what follows, I will tackle this issue head on and argue that (i) a new business model based on PDSes holds great promise in addressing the privacy (and other) challenges of Big Data, although this model inevitably invites debate over the '"propertization' of personal information; (ii) a propertized model is defensible as long as it provides the necessary safeguards for ensuring information privacy; and (iii) because PDSes satisfy the requirements of this privacy-protective model, the EU should encourage their development by providing regulatory incentives for firms that adopt them.

Does consumer empowerment address Big Data's privacy challenges?

Consumer empowerment

A growing number of commentators and activists, mainly in the USA and the UK, are engaged in describing and fostering a new business model premised on consumer empowerment. This represents a fundamental shift in the management of personal data 'from a world where organizations gather, collect and use information about their customers for their own purposes, to one where individuals manage their own information for their own purposes—and share some of this information with providers for joint benefits'.[53] Doc Searls, in his book *The Intention Economy: When Customers Take Charge*, makes the case for a new commercial order in which customers are emancipated from systems built to control them and become 'free and independent actors in the marketplace, equipped to tell vendors what they want' and how, where, and when they want it and at what price.[54] Searls describes a new category of tools for expressing and signalling customer intent, which he calls VRM (for vendor relations management). These VRM tools work as 'the demand-side counterpart' of vendors' CRM (customer relationship management) systems, which Searls derides as bastions

48 Bygrave (n 42), at 21 (recommending the elaboration in concrete contexts of the key principle implicit in Article 15, namely, that 'fully automated assessments of a person's character should not form the sole basis of decisions that significantly impinge upon the person's interests'); Zarsky (n 16), at 53–5 (proposing a public awareness campaign to help mitigate the problems associated with data mining); Zarsky (n 17), at 51–5 (proposing that firms provide notice of tailoring and the information they relied upon in tailoring content or ads and promoting secondary markets to undermine price discrimination); Hildebrandt (n 43), at 249 (recommending 'an *effective* right of access to profiles that match with one's data and are used to categorize one, including the consequences this may have'); Hildebrandt (n 24), at 53 (recommending a new emphasis on 'transparency by design' in the form of 'effective transparency enhancing tools (TETs) that allow citizens to anticipate how they will be profiled and which consequences this may entail').

49 Tene and Polonetsky (n 21), at 24.

50 Ibid, at 29.

51 Ibid.

52 This argument bears an obvious resemblance to the rationale for data portability. However, Tene and Polonetsky reject data portability as going 'too far' and even suggest that it might be anticompetitive or stifle innovation. This part of their argument is puzzling and not very persuasive.

53 Ctrl-Shift, 'The New Personal Data Landscape', 22 November 2011 available at <http://ctrl-shift.co.uk/about_us/news/2011/11/22/the-new-personal-data-landscape> accessed 17 December 2012.

54 Doc Searls, *The Intention Economy: When Customers Take Charge* (Boston: Harvard Business Review Press, 2012).

of guesswork and waste. Instead of a data gathering industry in which a new breed of hidden persuaders surreptitiously track and monitor user behaviour and preferences, and aggregate and exchange data for the purpose of forming educated guesses about what users want, all in the name of selling targetted ads that just might result in consumer purchases, Searls' vision is one in which 'demand finds supply'. In other words, individuals would rely on new VRM tools to express what kinds of information they are willing to release, to whom, and under what conditions. VRM further assumes that vendors are ready to receive and bid on such 'personal RFPs' securely and automatically.[55]

Nor is Searls a lonely prophet. In 2010, the World Economic Forum—with contributions from academics, privacy groups, and experts at major US and European IT firms—launched a project entitled 'Rethinking Personal Data'.[56] Like Searls, this group sees personal data as 'generating a new wave of opportunity for economic and societal value creation' but only if various stakeholders succeed in establishing a 'balanced personal data ecosystem'. The pivot point of this ecosystem is the concept of 'user-centricity', which seeks to integrate diverse types of personal data while putting end users at the centre of data collection and use, subject to a set of global data principles that include transparency, trust, control, and value creation.

This emphasis on consumer empowerment as the heart of a new business model presupposes that individuals maintain control over the creation and sharing of their personal data. This in turn depends on the availability of PDSes, which provide both a secure data store for a wide variety of personal information (including official records like birth and marriage certificates, licences, and passports, transaction records, online profiles, and social media content, and user names and passwords) as well as a new class of user-driven services ranging from personal RFPs as described above, to more participatory forms of healthcare, to 'FixMyStreet' and similar grass-roots citizenship efforts. Searls opines that most users will turn to 'fourth parties' for assistance in maintaining their PDSes, that is, to agents whose interests are strictly aligned with those of individual end users and who serve them in a fiduciary role.[57]

Based on the work of Searls and others, PDSes have eight main elements:[58]

1. individuals as the centre of personal data collection, management and use.[59]
2. selective disclosure, ie, the ability of customers to share their data selectively, without disclosing more personal data than they wish to.
3. control over the purpose and duration of primary and secondary uses. Control may be achieved by 'owner data agreements'[60] and/or by technical means such as DRM or meta-tagging (which are discussed at length below).
4. signalling, ie, a means for individuals to express demand for goods or services in open markets, not tied to any single organization.
5. identity management, which handle tasks such as the authentication and use of multiple identifiers while preventing correlation unless permitted by the user;

55 For an alternative depiction of VRM in terms of 'user driven services' in which 'users start each interaction, manage the flow of the experience, and control what and how data is captured, used and propagated', see Joe Andrieu, Introducing User Driven Services, joeandrieu.com, April 26, 2009 (series of ten blog posts) <http://blog.joeandrieu.com/2009/04/26/introducing-user-driven-services/> accessed 3 September 2012.

56 World Economic Forum, 'Rethinking Personal Data' (2010) and 'Personal Data: The Emergence of a New Asset Class' (2011), both available at <http://www.weforum.org/issues/rethinking-personal-data> accessed 17 December 2012.

57 Searls (n 54), at 177–79. Also see Jerry Kang et al., 'Self-Surveillance Privacy' (2012) 97 Iowa L. Rev. 809 (describing the need to house vital signs and other 'self-measurement' data in data 'vaults' managed by personal data 'guardians' who owe fiduciary duties to their individual clients including duties of care, confidentiality, and loyalty). Ideally, guardians would be treated as professionals subject to conflict of interest rules. For example, they would be prohibited from exploiting their access to an individual's data by engaging in data mining in exchange for free services. For a more skeptical review of personal data stores, infomediaries, and VRM systems, see Arvind Narayanan et al., 'A Critical Look at Decentralized Personal Data Architectures' (2012) <http://arxiv.org/abs/1202.4503> accessed 17 December 2012.

58 This analysis relies on Searls (n 54); Ctrl-Shift (n 53); Andrieu (n 55); World Economic Forum (n 56); and Mydex, 'The Case for Personal Information Empowerment: The Rise of the Personal Data Store', September 2010 <http://mydex.org/wp-content/uploads/2010/09/The-Case-for-Personal-Information-Empowerment-The-rise-of-the-personal-data-store-A-Mydex-White-paper-September-2010-Final-web.pdf> accessed 17 December 2012.

59 A 2011 paper by the World Economic Forum (n 56) offers an interesting definition of personal data as encompassing 'volunteered data'—created and explicitly shared by individuals (eg a social networking profile); 'observed data'—captured by recording the actions of individuals (eg location data); and 'inferred data'—data about individuals based on analysis of volunteered or observed data (eg a credit score). Presumably, PDSes would include all three types of personal data. Whether an individual can or should control observed and inferred data is a difficult question. It raises both architectural issues (for example, how and when do observed and inferred data become part of a PDS) and free speech issues (for example, should individuals have a veto power over discreditable information about themselves)? This requires a longer and more nuanced discussion than is possible here.

60 For example, Personal.com is a start-up enabling individuals to own, control access to, and benefit from their personal data. See Meet the Owner Data Agreement, available at <https://www.personal.com/legal-protection> accessed 3 September 2012).

6. security, ie, a very high level of security as more fully described below;
7. data-portability, ie, the ability to move all of one's data from one provider to another using standard data formats and interface protocols; and
8. accountability and enforcement, ie, accountability for protecting and securing personal data in accordance with the rights and permissions established by agreement and/or enforced by tagging mechanisms; and enforcement under self-regulatory guidelines and legal mandates, both backed by comprehensive auditing.

What is immediately striking about these eight elements is how well they comport with the core EU data protection principles. To begin with, the entire scheme ensures a much higher degree of transparency than would be achievable with the opaque data stores maintained by businesses or third parties. Elements 1–3 address purpose specification as well as collection and use limitations. Element 1 strongly supports data quality, while elements 6 and 8 match up, respectively, with data security and accountability. Moreover, PDSes treat FIPs not as externally imposed constraints that must be balanced against business objectives, but as a set of organizing principles, which guide business objectives and thereby set design goals from the outset. The next question, then, is whether PDSes also address the privacy challenges of Big Data.

PDSs and Big Data

PDSes face many obstacles—ranging from the technical (adequate security, establishing a new permission model based on meta-tagging, preventing re-identification); to the legal (establishing a legal framework supporting propertized personal information, developing a co-regulatory approach that incentivizes, rather than penalizes, new business models, harmonizing international legal rules); to a variety of business and social tasks implicit in creating a new ecosystem.[61] In assessing this new business model as a possible solution to the privacy challenges of Big Data, however, we limit ourselves to three central concerns: technical feasibility, intellectual coherence, and the existence of business incentives.

Technical feasibility

In order to be technically feasible, PDSes must meet two main requirements: security at a very high level and the ability to enforce privacy rights by 'tagging' every unit of personal data with metadata describing privacy-related requirements and preferences. One or more PDS will store individuals' most personal and sensitive data in a comprehensive fashion. As such they will be prime targets for both internal and external attacks. Thus, they are inconceivable without the highest level of security. They will need to be designed with at least the following requirements in mind: (i) all personal data must be encrypted both in storage and during transmission; (ii) all encryption keys must be stored outside the PDS; (iii) all metadata must be encrypted and digitally signed; (iv) all individuals who access a PDS must be authenticated by multi-factor authentication and authorized to perform various actions; and (v) all PDSes must ensure accountability by using secure audit mechanisms.[62]

As to rights enforcement via tagging, over a dozen years ago, Professor Jonathan Zittrain argued that digital rights management (DRM) systems might be adopted for privacy purposes. According to Zittrain, these trusted systems 'structure "rights" into a calculable framework that is then automatically enforced by the technology'.[63] His article describes how a hypothetical trusted system might handle the core aspects of health privacy rights. Zittrain's key insight is that DRM permits individuals to share personal data while maintaining control. Korba and Kenny have since proposed a 'privacy rights management' (PRM) system to support privacy principles derived from the DPD. Under this model, the data controller acts as the enforcer of the usage requirements for personal data and manages the collection, storage, and processing of personal data from the data subject.[64] Their approach falls short of the 'control shift' envisaged by the development of PDSes in which individuals act as data managers. What's missing from their work is a stronger notion of 'permissible uses' determined by the data subject and enforced by meta-data associated with individual elements of personal data, using a rights expression language such as XrML or ODRL.[65]

Perhaps the most sophisticated effort to date at defining a DRM-based approach to privacy is the PCAST

61 World Economic Forum, 'Rethinking Personal Information: Workshop Pre-Read' (2010).

62 See PITAC (n 66).

63 Jonathan Zittrain, 'What the Publisher Can Teach the Patient: Intellectual Property and Privacy in an Era of Trusted Privication' (2000) 52 Stanford L. Rev. 1201, 1212.

64 Larry Korba and Steve Kenny, 'Towards Meeting the Privacy Challenge: Adapting DRM' DRM Workshop (2002) available at <http://crypto.stanford.edu/DRM2002/KorbaKennyDRM20021.pdf> accessed 17 December 2012.

65 See Viktor Mayer-Schönberger, 'Beyond Copyright: Managing Information Rights with DRM' (2006) 84 Denver Univ. Law Review 181, 186–9.

report on IT health systems, which describes a universal exchange of health data among electronic health systems using 'meta-data tagged' data elements.[66] According to the vision laid out in this report, 'the metadata attached to each of these data elements ... would include (i) enough identifying information about the patient to allow the data to be located (not necessarily a universal patient identifier), (ii) privacy protection information—who may access the [patient information], either identified or de-identified, and for what purposes, and (iii) the provenance of the data—the date, time, type of equipment used, personnel (physician, nurse, or technician), and so forth.'[67] The report further envisions the creation of a national infrastructure providing 'data-element access services' including those associated with crawling, indexing, and searching the data being exchanged, both for treatment and research purposes. Through the use of metadata and related security components, this universal exchange system would have a number of interesting properties. First, the access services would not have the right to use—or even necessarily view—the underlying data. Rather, the system would be structured 'to expose only those data elements authorized by privacy rules and policies ... and only to authenticated, authorized users', thereby allowing patients to opt in or out of these access services.[68] Second, the data required by any access service may be gathered from various servers and aggregated, analysed, and presented in real time (akin to browsers assembling the elements of a Web page on a just-in-time basis). As a result, the health IT ecosystem would remain highly decentralized.[69] Third, this ecosystem avoids the use of uniform patient identifiers. Finally, tagged data elements would enable effective implementation of privacy rules and policies. As opposed to the largely empty ritual of privacy policies or the all-or-nothing choices typical of most Web sites today, a tagging system allows for fine-grained individual privacy preferences.

This vision of an IT health ecosystem based on tagged data elements remains just that—a vision. Nor is it obvious that what might be achieved in the highly regulated and partially closed health IT setting is readily transferable to the weakly regulated and open-ended commercial Internet setting. (Or that the costs would be justified absent a viable business model.) Moreover, the music industry has met with only limited success in relying on DRM to prevent illegal use and sharing of content. Hence it remains to be seen whether appropriate norms and incentives could be designed that would ensure greater success for DRM in controlling the unwanted sharing of health data or any other personal data. Rather, the point being made here is that the relevant technology is available and hence such systems pass the technical feasibility test.

Intellectual coherence

PDS proponents inevitably speak of individual control *and ownership* of personal data. The adoption of DRM technology for data protection purposes only intensifies this association between personal data and property-related actions like trading, exchanging, or selling data.[70] Although proponents of PDSes are well aware of the enormous challenges they face, here we focus exclusively on a single issue, namely, whether it is possible to overcome longstanding objections to a 'propertized' conception of information privacy and in this way ensure that this new business model is intellectually coherent.

As Professor Paul Schwartz points out, legal scholars who have advocated 'propertization of personal information' have often shown insufficient sensitivity to privacy concerns, while those who oppose it have generally advocated an outright ban on data trade, rather than restrictions on transferability. In a groundbreaking article, however, Schwartz offers 'a model for propertization of personal data that will fully safeguard information privacy'.[71] Schwartz identifies three main concerns with a property-based theory. First, propertization will exacerbate privacy market failures. Schwartz briefly summarizes the longstanding view of privacy scholars that existing markets for privacy are dysfunctional: 'Because the gatherers have greater power to set the terms of the bargain and to shape the playing field that guides individual decisions, at the end of the day negotiations in the privacy market may fall short.'[72] Second, propertization will neglect important social values that information privacy should advance. Schwartz sees information as a public good (like clean air or national defence). He refers to the public good at

66 President's Council of Advisors on Science and Technology (PCAST), 'Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward' (2010) <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf> accessed 17 December 2012.

67 Ibid, at 41.

68 Ibid, at 42.

69 Ibid. A centralized health IT ecosystem or PDS would be a single point of failure, which is always problematic from both a security and privacy standpoint.

70 See Mayer-Schönberger (n 65), at 194.

71 Paul M Schwartz, 'Property, Privacy, and Personal Data' (2004) 117 Harvard Law Review 2055, 2058.

72 Ibid, at 2081–82.

stake as 'the privacy commons—a space for anonymous and semi-anonymous interactions'.[73] Schwartz therefore rehearses traditional public good arguments that cast doubt on the propertization of information and calls for limits on the propertization of personal information 'to the extent that it undermines the privacy commons'.[74] Third, propertization invites or even entails free alienability of personal data, which is highly problematic for two reasons: secondary use of personal data and the difficulty of estimating the appropriate price for such secondary uses.[75]

Schwartz does not rebut these concerns so much as overcome them by offering a model of propertized personal information conditioned on privacy safeguards. After briefly summarizing his model, I will argue that PDSes with tagged elements fully instantiate it. The five elements of Schwartz's model are as follows:

- *Limitations on an individual's right to alienate personal information.* To limit the market failures associated with 'one-shot' permissions for data trade, in which users 'have only a single chance to negotiate future uses of their information', and unsurprisingly trade away too much of their propertized personal information, Schwartz proposes a combined use-transfer restriction in which property is an interest that 'runs with the asset'. Thus, use-transferability restrictions 'follow the personal information through downstream transfers'.[76]
- *Default rules that force disclosure of the terms of trade.* Noting that defaults promote individual choice, Schwartz favours opt-in defaults, because they reduce information asymmetry problems by forcing 'the disclosure of hidden information about data-processing practices'.[77] Schwartz proposes a model of 'hybrid alienability' that combines a use-transferability restriction with an opt-in default. 'In practice,' he writes, 'it would permit the *transfer* for an initial category of *use* of personal data, but only if the customer is granted an opportunity to block further transfer or use by unaffiliated entities.'[78]
- *A right of exit for participants in the market.* This right of exit may be thought of as an expanded form of consent in which individuals can not only refuse data trades upfront but exit from an agreement to trade (and thereby get out from a bad bargain); additionally, they may remove or disable tracking technologies, and not only exit but also re-enter data trades as they wish.[79]
- *Damages to deter market abuses.* Schwartz recommends an enforcement scheme premised on liquidated damages mainly because the higher damages typical of this approach encourage companies to keep their privacy promises and overcome collective action problems of consumers whose individual damages for privacy violations may be too low to bear the costs of litigation.[80]
- *Institutions to provide trading mechanisms (a 'market-making' function), to verify claims to propertized personal data (a verification function) and to police compliance with bargained-for terms and legal safeguards (an oversight function).* Schwartz argues that these institutions 'will assist the privacy market by ensuring that processes exist for the exchange of data and for the detection of violations of privacy promises'.[81] His preferred model of these institutions involves decentralization (ie, multiple small markets) and individual enforcement (via private rights of action) supplemented by the FTC's ongoing role in policing privacy promises. Most interestingly, he calls for the verification of propertized personal information through an association with 'nonpersonal metadata'.[82]

Now comes the ultimate question: do PDSes satisfy all five elements of Schwartz's model of hybrid propertization? The short answer is 'yes', at least for all intents and purposes. First, PDSes are premised on use-transferability restrictions, and PDSes with tagged data permit fine-grained expression and enforcement of such restrictions. Second, these systems not only combine use-transferability restrictions with an opt-in default but also offer users a tagging mechanism for enforcing initial and subsequent choices in a persistent and reliable manner. Third, PDSes readily allow users to opt-in or out of various services at any time, thereby enabling a right of exit. Fourth, while liquidated damages are orthogonal to PDSes, they may also be unnecessary if the system works as designed. That said, one could readily imagine legislation that would impose liquidated damages on PDSes, service providers, third and fourth parties, or even individuals, who engage in malicious or abusive acts or negligently fail to carry out their obligations. Finally, the vision sketched out above (eg, in the

73 Ibid, at 2089.
74 Ibid, at 2088.
75 Ibid, at 2091.
76 Ibid, at 2098.
77 Ibid, at 2100.
78 Ibid, at 2098.
79 Ibid, at 2106–07.
80 Ibid, at 2107–09.
81 Ibid, at 2100.
82 Ibid, at 2112.

PCAST report) incorporates institutional elements that address the three functions Schwartz identifies (market-making, verification, and oversight) in a decentralized fashion. Again, one could imagine the enactment of additional legal safeguards to supplement the technical enforcement provided by tagging even though legal rules remain external to the vision sketched out above.

Business incentives

Searls' vision of an intention economy might strike some readers as fanciful. Given the importance of data mining to online commerce, the size and power of the online advertising industry, and the seemingly entrenched status of the ad-funded business model, why should businesses embrace a new model that requires them to shift control over data to individual users? Three reasons may be adduced. First, PDSes serve a compelling need—they help individuals organize and manage their daily lives and give them tools for realizing the inherent value of their own data. Thus, they are both convenient and a source of insight (via a new class of apps for monitoring and analysing one's own behaviour). As such, PDSes not only create new sources of economic and social value, but present an intriguing business opportunity, especially for new entrants who lack access to the Big Data silos of incumbents like Acxiom, Google, Facebook, etc. Second, and quite apart from this competitive dynamic of new entrants vs. incumbents, PDSes hold out the promise of supplying firms with better data. Data mining and ad targeting are based on guesswork, whereas in the intention economy, potential customers would knowingly and intentionally reveal data that are likely more to be detailed, accurate, complete, and up-to-date than any inferred data. Finally, PDSes may enhance customer trust and quite possibly would lower compliance costs for firms that rely on data in PDSes as opposed to collecting and storing data in their own data stores.

In sum, this section has demonstrated that PDSes not only strongly support the core FIPs but also are technically feasible, intellectually coherent, and potentially attractive to business. That said, PDSes are not a privacy panacea and their strengths and weaknesses must be assessed in specific settings. In particular, more research is needed to determine what other technical and legal safeguards are needed to complement PDSes in a variety of scenarios where their value may be limited.

Recommendations and conclusion

In light of the foregoing critique of the Regulation and the apparent benefits of PDSes in addressing Big Data privacy concerns, what other steps should EU privacy officials consider? In particular, what might they do to foster this new business model that so strongly supports individual empowerment and thereby may accomplish by other means many of the same goals that others hope to achieve solely through regulation?

First, EU officials should support private sector efforts to stand up PDSes by funding relevant research projects. These might include development of new APIs and rules languages; review and endorsement of data ownership agreements and of best practices for security and meta-tagging; and/or development of model laws describing the fiduciary duties (care, confidentiality, loyalty) of data custodians. Second, they should investigate whether the use of meta-tagging systems to enforce privacy rules, policies, and preferences requires new laws making it an actionable offence to tamper with or act in contravention of securely tagged data elements. Third, and most importantly, they should encourage firms to experiment with PDSes by offering them certain regulatory incentives. For example, firms running PDSes that combine all eight elements as described above might enjoy regulatory 'carrots' (such as relaxation of some of the enhanced responsibilities of data controllers),[83] together with avoidance of regulatory 'sticks' (such as the very high penalties imposed on firms that that fail to meet these requirements).[84] It is important to emphasize that under such a co-regulatory approach, these incentives do not involve any derogation from fundamental rights. Rather, the logic of rewarding firms that voluntarily pursue this new business model premised on PDSes is simply this: since the primary purpose of the Regulation's enhanced responsibilities (and new penalties) is to ensure compliance with core data protection principles, and since the principles under which PDSes operate inherently respect the rights of data subjects, it follows that burdening these firms with additional obligations and the risk of severe fines is counterproductive.[85]

83 By 'enhanced responsibilities' I have in mind the accountability provision (Article 22), data protection by design and default (Article 23), documentation (Article 28), impact assessments (Article 33), prior authorization or consultation (Article 34), and designation of a data protection officer (Article 35), but *not* data security (Article 30), which remains one of the core data protection principles.

84 Enterprises that intentionally or negligently fail to comply with the documentation provision face fines of up to €500,000 or 1% of annual worldwide turnover, Article 79(5)(f) of the Regulation, while those that fail to comply with the remaining obligations identified in note 83 are subject to double these fines, Article 79(6)(e), (i) and (j).

85 For a similar analysis, see Hans-Bredow Institute, 'Final Report: Study on Co-Regulation Measures in the Media Sector, 119–20 (2006), <http://ec.europa.eu/avpolicy/docs/library/studies/coregul/final_rep_en.pdf> accessed 17 December 2012.

Nor is it necessary for the Commission to invent a new regulatory vehicle for conducting this experiment in co-regulation. Both the DPD (Article 27) and the Regulation (Article 38) allow industry sectors and other 'categories of controllers' to submit codes of conduct to member states or to the Commission for approval. Because approval turns on a finding that the code is 'in compliance' with the DPD or the Regulation, approved codes of conduct arguably serve as a legal safe harbour for firms that comply with them.[86] Moreover, Article 38(4) allows the Commission to adopt 'implementing acts' for deciding that the codes of conduct have community-wide validity. Thus, the Commission has the power to begin experimenting with an incentive-driven, co-regulatory approach in which firms draft and submit codes of conduct for review, and regulators remain responsible for code approval, oversight, and enforcement.[87]

There are several additional steps that might be taken by the Council of the European Union or the European Parliament to ensure that such co-regulatory experiments are fully supported in the Regulation. In particular, they should consider modifying the proposed Regulation by:

- adding language to Article 38 (i) clarifying that companies that adhere to an approved code of conduct will be treated as having complied with the Regulation for purposes of any enforcement action unless the code was approved on the basis of false or incomplete information; and (ii) granting the Commission the power to negotiate tailored requirements for certain industry sectors by relaxing or granting exemptions from the new responsibilities described in Articles 22–23, 28, or 33–34; and
- adding language to Article 79 reducing the fines imposed on firms that violate Articles 22–23, 28, or 33–34 but otherwise comply with an approved code of conduct and/or modifying Article 79(2) by adding 'the code of conduct implemented pursuant to Article 38' to the list of discretionary factors a supervisory authority may consider in assessing fines.

This paper has argued that privacy laws such as the DPD have failed to keep pace with technological changes or to limit the access of both the public and private sectors to large and diverse datasets for analytic purposes. Big Data only amplifies these trends. The proposed Regulation seeks to address these changes but does not fully grapple with the challenges posed by Big Data. But the EU need not rely solely on regulatory reform to achieve its data protection goals. Rather, as argued above, it can and should take advantage of emerging business models in which firms decide for sound business reasons to empower consumers and enhance individual control over personal data. One way to do this is by experimenting with a more flexible approach to regulation through the creative use of codes of conduct.

Of course, it is too soon to say whether firms will embrace these new business models, especially if they entail satisfying the stringent security and privacy requirements identified above. Nor is it clear that consumers would be better off if PDSes become prevalent—perhaps data-driven businesses will find ways to circumvent these protections. Further study is required, including very detailed investigations of how PDSes would operate in specific sectors (eg, e-commerce, health care, smart environments, and social networking) together with an analysis of problematic issues such as how ad-funded businesses would respond, whether the new business model would replace or merely coexist with the advertising model, and what the transition path might look like for companies wishing to make the change. And yet if the critique offered above is well founded, this strongly suggests that regulatory reform alone is not enough and needs help from other quarters. Thus, new and innovative business models that shift control from companies to individuals, subject to the privacy conditions set out above, may be worth exploring.

doi:10.1093/idpl/ips036

86 Dennis D Hirsch, 'Dutch treat? Collaborative Dutch privacy regulation and the lessons it holds for U.S. privacy law' (forthcoming 2013) Michigan State Law Review (discussing codes of conduct under Dutch law).

87 For a detailed analysis of co-regulation as it applies to privacy, see Ira S Rubinstein, 'Privacy and Regulatory Innovation: Moving Beyond Voluntary Codes' (2011) 6 I/S: A Journal of Law and Policy for the Information Society 555 (distinguishing co-regulation from self-regulation).