THE COLUMBIA

SCIENCE & TECHNOLOGY

LAW REVIEW

VOL. XVII STLR.ORG SPRING 2016

ARTICLE

# REGIME CHANGE? ENABLING BIG DATA THROUGH EUROPE'S NEW DATA PROTECTION REGULATION[†]

Viktor Mayer-Schönberger[*] & Yann Padova[◊]

*The European Union has just passed the most comprehensive overhaul of its privacy laws in two decades. The so-called General Data Protection Regulation (GDPR) brings numerous changes. In this Article we look at how the GDPR will impact, enable or restrict the use of Big Data in Europe. We look at three distinct privacy aspects of Big Data: while collection sees a tightening of regulation through the GDPR, the principles of purpose limitation and retention minimization offer a mixed view, including a rather surprising avenue that could permit Big Data applications at a much larger scale than today. We conclude in evaluating whether this exception, that provides some room for manoeuver for member states, will be sufficient for Big Data to flourish, and what issues remain to be addressed*

[†] This Article may be cited as http://www.stlr.org/cite.cgi?volume=17&article=SchonbergerPadova.

[*] Professor of Internet Governance and Regulation, University of Oxford.

[◊] Former Secretary General of the French Data Protection Authority (CNIL), now Commissioner with the French Energy Regulator (CRE).

## INTRODUCTION

Some have described Big Data—using comprehensive data to gain novel insights into how the world works that would not have been attainable using small amounts of data—as bringing about fundamental changes in how humans live, work and think.[1] It is seen as the next "big thing" that will generate enormous value and help those that understand its power to reap huge benefits.

By the same token, for Big Data to reach its potential, data needs to be gathered at an unprecedented scale whenever possible, and reused for different purposes over and over again. This is already happening. With ten billion devices connected to the Internet today and between thirty and eighty billion in 2020, the total amount of data in the world is predicted to double every twenty-four months.[2] Big Data's raw material is not going to be scarce. This puts Big Data on a direct collision course with the core principles of existing data protection laws. Lawmakers around the globe are struggling to find a new balance between the need to protect the information privacy of individuals against the demand to utilize the latent value of data.

Moreover, finding this balance does not take place in a vacuum. Nations are vying for high tech activity through legal and regulatory frameworks. Time is scarce as technology companies are innovating at an astounding pace, challenging legal systems to keep up and remain competitive.

In this Article, we examine how the European Union has recently confronted these challenges as part of the far-reaching overhaul of its privacy legislation. We suggest that the new European Union General Data Protection Regulation (GDPR),[3] agreed to in December 2015 after years of tense negotiations among the political stakeholders in Brussels, opens up some

[1] One of us has even suggested that Big Data may spur a "revolution." *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH N. CUKIER, BIG DATA: A REVOLUTION TRANSFORMING HOW WE LIVE, WORK, AND THINK (2013).

[2] INSTITUT MONTAIGNE, BIG DATA ET OBJETS CONNECTÉS: FAIRE DE LA FRANCE UN CHAMPION DE LA REVOLUTION NUMÉRIQUE 15–16 (2015).

[3] At the time of writing, the GDPR had not been formally enacted, but the relevant European Union institutions have reached formal agreement on the exact text of the regulation. This Article is based on that formally agreed text. *See* Council of the European Union, 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) No. 15039/15 of 15 December 2015 [hereinafter GDPR].

concrete doors for Big Data to flourish although it falls short of abandoning traditional data protection principles altogether. The new data protection setup is far from being simple or transparent, leaves much clarity to be desired, and gives significant leeway to Member States, but it is a remarkable step towards enabling Big Data in Europe.

In the Article's first part, we highlight the key features of Big Data.[4] In the second part,[5] we describe how the existing 1995 European Union Data Protection Directive (DPD)[6] severely curtails the use of Big Data. In the third part,[7] we analyze the main pathways permitting Big Data under the new GDPR, and, in the fourth part, the GDPR's limitations and contradictions.[8]

## I. BIG DATA AND ITS DEFINING QUALITIES

Big Data is often described as a new technology or a set of new technical tools that aids in the collection and mathematical analysis of data, using traditional statistical methods as well as more innovative analytical tools. Yet such a view of Big Data fails to capture what makes Big Data so special as well as powerful. Big Data actually opens up a new perspective on reality. Of course, humans have always made sense of the world around them by observing it. Collecting and analyzing data is foundational to how we have tried to understand the world. But in analog times, utilizing data was hard, time-consuming and costly. Because of these constraints in working with data, the processes, structures and institutions of discovery and sense-making were designed to use as little data as absolutely necessary, essentially internalizing the notion that usable data was scarce.

The technological changes brought about by digitization have reduced dramatically the time and cost required to gather and analyze data. Where in the past at best a small sample of data could be collected and examined, in the future, vastly more data, not just in absolute terms but *relative* to the phenomenon to be studied, can and will be utilized.[9] This offers vastly more detail and an unprecedented comprehensiveness, enabling us to zoom in on data much more freely.[10] Such comprehensive use of data reduces

[4] *See infra* Section I.

[5] *See infra* Section II.

[6] Council Directive 95/46, 1995 O.J. (L 281) [hereinafter DPD].

[7] *See infra* Section III.

[8] *See infra* Section IV.

[9] *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 28.

[10] *Id.* at 26–31.

some (but certainly not all) of the possibilities of bias and error, because limited data samples no longer have to be extrapolated to the whole.[11] More importantly, comprehensive data can be used not just to answer concrete questions, but to stimulate new ones not yet thought about.[12]

A case in point is Big Data analysis undertaken by the creators of the foreign-language learning app "Duolingo."[13] Duolingo—used by tens of millions of individuals worldwide—captures their responses, including erroneous ones, as they learn a foreign language. When the Duolingo engineers looked at the response data, they discovered a surprising pattern among native Spanish speakers aiming to learn English: they made steady progress until they reached a particular lesson, but after that often seemed confused. Resequencing that particular lesson to sometime later in the program would significantly improve success rates. It wasn't something that the engineers knew or thought of when implementing data collection in the app, nor was it a concrete hypothesis they already had and wanted tested. Rather, the pattern in the data suggested a novel hypothesis that not only led to a discovery about how Spanish speakers best learn English, but ended up significantly improving the product.

This quest to identify novel patterns in data is what Big Data experts call "letting the data speak."[14] At least to an extent, it reverses the direction of discovery, using data to foster hypotheses rather than "prove" existing hypotheses.[15] As the analysis of comprehensive data offers valuable new perspectives on the world, it greatly facilitates human discovery, which in turn may lead to economically valuable innovations.

In the future, the source of discovery will shift, at least in part, to data and its analysis. As a result, the value of data will change. If in the past, the value of data was captured by collecting and using it once for a concrete purpose, with Big Data the latent value of data is unclear at the time of collection and can only be fully

[11] *Id.* at 30–31.

[12] *See* Viktor Mayer-Schönberger, *Big Data for cardiology: novel discovery?*, EUROPEAN HEART J., Dec. 24, 2015, *available at* http://eurheartj.oxfordjournals.org/content/early/2015/12/24/eurheartj.ehv648 ("researchers have recently employed a modified approach whereby the hypotheses are algorithmically generated and then tested against data").

[13] *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, LEARNING WITH BIG DATA 9–11 (2014).

[14] MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 19 (quoting Jeff Jonas, IBM engineer and Big Data pioneer).

[15] *See* MAYER-SCHÖNBERGER, *supra* note 1, at 55.

reaped as the data is being reused over and over again for different purposes. Moreover, the value of data can be greatly enhanced not only by having and analyzing more of it, but by combining it with other data sources. It is like a single puzzle piece that taken by itself offers little value, but when combined with others to complete an image is turned into something precious.

The shift in the value of data creates a very strong economic incentive in how data is being handled: it will be collected whenever there is a possibility to collect even though no concrete use case is evident; collection is opportunistic rather than purposeful. Similarly, there is a strong economic incentive to keep the data for as long as possible, and much beyond the initial use of it, to reuse it repeatedly as well as to combine it with other data.

## II. BIG DATA UNDER THE DPD: THE CLASSICAL APPROACH

The European Union DPD, enacted more than twenty years ago in 1995, was the EU's first general data protection act covering both the private and the public sectors.[16] It took years of negotiations, but in the end it incorporated the core principles[17] of informational data protection that had evolved in the decades before and had been incorporated in national legislation and international agreements. As a directive, it was not directly applicable, but had to be transposed into national laws.[18] However, its text left national legislators relatively little flexibility. The explicitly stated goal was to guarantee a high level of information privacy throughout Europe that would enable the "free flow" of data within the EU.[19]

[16] Some processing of personal data is specifically excluded, such as processing for public security, defense and state security. *See* DPD, *supra* note 6, at art. 3.2.

[17] *See, e.g.*, SEC'Y'S ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS (1973), https://perma.cc/9JBY-AKXC (recommending the adoption of a Code of Fair Information Practice) [hereinafter FIPP]; ORG. FOR ECON. COOPERATION AND DEV., GUIDELINES ON THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA, (1980), https://perma.cc/U4RY-JF6W [hereinafter OECD Guidelines].

[18] *See* DPD, *supra* note 6, at arts. 32, 34.

[19] This is already obvious in the official title "Directive . . . on the protection of individuals with regard to the processing of personal data and *on the free movement of such data*" (emphasis added), *supra* note 6; it is also explicitly stated in Article 1.2 (Object of the Directive): "Member States shall neither restrict nor prohibit the free flow of personal data between Member

The directive does not predate the Internet, but it was drafted when the Internet was still little more than a niche network, connecting mainframes, minicomputers and a small but growing number of PCs through slow dialup connections. Smartphones did not exist, storage space was measured in megabytes, e-commerce was just being born, and widespread social media was science fiction. Unsurprisingly, the directive reflects a "small data" world in which data collection, storage and processing is still comparatively expensive and thus undertaken sparingly.

Nowhere is this more obvious than in the directive's purpose limitation principle.[20] It states that the processing of personal data can only be undertaken for a concrete purpose, and once this purpose has been fulfilled, personal data has to be discarded. This reflects the principle of data minimization.[21] Repurposing data for novel purposes is explicitly prohibited unless the further processing is not "incompatible." [22]

For the processing of personal data to be lawful, the directive further requires that the individual the personal data pertains to (called the "data subject") either has given her consent,[23] the processing is necessary for the "performance of a contract" with the data subject,[24] or to comply with a legal obligation of the data controller.[25] Processing is also permitted if it is in the "vital interests

States for reasons connected with the protection afforded under paragraph 1.", *supra* note 6, at art 1.2.

[20] DPD, *supra* note 6, at art. 6.1(b).

[21] *Id.* at art. 6.1(e).

[22] *Id.* at art. 6.1(b). The assessment of compatibility has long been a difficult issue for data controllers. The Article 29 WP issued some guidance in 2013, *Opinion 03/2013 on purpose limitation*, Article 29 Data Protection Working Party, No. 00569/13/EN, (2013), [hereinafter Article 29 WP], and provided the following criteria in order to assess the compatibility of the further processing: "Further processing for a different purpose does not necessarily mean that it is incompatible: compatibility needs to be assessed on a case-by-case basis. A substantive compatibility assessment requires an assessment of all relevant circumstances. In particular, account should be taken of the following key factors: the relationship between the purposes for which the personal data have been collected and the purposes of further processing; the context in which the personal data have been collected and the reasonable expectations of the data subjects as to their further use; the nature of the personal data and the impact of the further processing on the data subjects; the safeguards adopted by the controller to ensure fair processing and to prevent any undue impact on the data subjects."

[23] DPD, *supra* note 6, at art. 7(a).

[24] *Id.* at art. 7(b).

[25] *Id.* at art. 7(c).

of the data subject,"[26] "carried out in the public interest or in the exercise of official authority,"[27] or fulfills the legitimate interest of the data controller, which is not overridden by the rights of the data subject.[28]

Quite obviously, these rules make it difficult for Big Data activities with personal data to take place in Europe. Many companies and organizations processing personal data have reacted by asking individuals to consent to the use of their personal data for very broadly defined purposes.[29] Arguably, this permits them to hold on to personal data—even after the initial purpose has been fulfilled—and to repurpose personal data as long as they can show that there is another purpose covered by the consent that they are in the process of using the data for. It does not, however, solve the challenge of getting every individual to accept a broadly phrased consent agreement in the first place. Nor does it permit consent forms that are so broad that they are seen as insufficiently "specified and explicit."[30]

There are two less obvious strategies to getting around the rather restrictive rules in the directive. The first is to "clean" the data so that it no longer contains any personally identifiable pieces of data. Then the directive no longer applies, as it is only focused on "personal data," defined by the directive as "any information relating to an identified or identifiable natural person."[31] The appeal of this de-identification approach, however, is limited in practice. As experts have shown, even data that at first glance

[26] *Id.* at art. 7(d).

[27] *Id.* at art. 7(e).

[28] *Id.* at art. 7(f).

[29] *See, e.g.*, the privacy notices and Terms of Service of Facebook: *Statement of Rights and Responsibilities,* FACEBOOK (Jan. 30, 2015), https://www.facebook.com/legal/terms; *Data Policy*, FACEBOOK (Jan. 30, 2015), https://www.facebook.com/about/privacy/; Amazon: *AWS Site Terms*, AMAZON (Dec. 23, 2011), https://aws.amazon.com/terms/; *AWS Privacy*, AMAZON (Oct. 7, 2015); and Google: *Google Terms of Service*, GOOGLE (Apr. 14, 2014), https://www.google.com/intl/en/policies/terms/; *Welcome to Google Privacy Policy* GOOGLE (Aug. 19, 2015), https://www.google.com/intl/en/policies/privacy/.

[30] One example of joint enforcement actions by European data protection agencies against a data controller because of too imprecise a purpose specification and broad a privacy policy statement is the case against Google and its privacy terms; *see, e.g.*, Commission Nationale de l'Informatique et des Libertés [CNIL] [French Data Protection Authority] Paris, Sanctions Committee, Jan. 3, 2014 [Deliberation No. 2013-420] (imposing a financial penalty against Google Inc. because of too imprecise a purpose specification and too broad a privacy policy statement).

[31] DPD, *supra* note 6, at art. 2(a).

seems unidentifiable (because, for instance, all seemingly identifying elements have been omitted), can become identifiable when combined with other data, and exposed to sophisticated statistical analysis.[32] Thus data would have to be purged of even the slightest personal identifiers in order to not be "personal."

The second strategy is to limit Big Data to statistical purposes, which constitute an explicitly permitted reuse of data, as long as Member States have put in place "appropriate safeguards."[33] Recital 29 further clarifies that such safeguards must "in particular rule out the use of the data in support of measures or decisions regarding any particular individual."[34] In short: statistical analysis is fine, as long as the member state has put in place safeguards that ensure that the analysis is not used in a concrete decision affecting a particular individual.

For example, a company cannot use the exemption for statistical analysis to utilize the personal data it has on its customers to devise a statistical model that predicts which customers are likely to defect to competitors and offer them special deals. It can, however, use the data to create a model that predicts the likely overall percentage of customer churn.

## III. BIG DATA UNDER THE GDRP

The European Union's new legal framework for data protection, the GDPR,[35] has been designed to address two overall weaknesses in the 1995 directive. First, technical advances have enabled data processing far beyond what was envisioned in 1995, creating new threats to individual privacy, but also new opportunities to reap value out of data. Thus, political decision-makers in the EU agreed that the directive needed some significant updating.[36]

Second, EU Member States not only transposed the directive's mandates slightly differently into their respective national laws, they also began to differ quite significantly in how they implemented and enforced these national laws. In its growing

[32] *See, e.g.*, Paul Ohm, *Broken Promises Of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716–27 (2010).

[33] DPD, *supra* note 6, at art. 20.

[34] *Id.* at Recital 29.

[35] GDPR, *supra* note 3.

[36] Unlike the DPD, which stressed the importance not only of data protection but also of free data flows throughout Europe, the GDPR's explicit aim is not to facilitate innovation in Europe. In fact, the word "innovation" is not mentioned once in the well over one hundred of the GDPR's Recitals.

number of rulings on data protection matters the European Court of Justice, the EU's highest court, has offered guidance on how to interpret and enforce the directive's mandates, but obviously can only do so in the context of actual controversies before it.[37] So a political desire emerged in the EU to reduce opportunities for forum shopping by Europeanizing the data protection framework, and centralizing some of the enforcement.[38] The redraft of the directive thus morphed into the drafting of a regulation, which is directly applicable law in all EU Member States.

The result is a bit of an unusual hybrid of old and new. It abolishes the need in many cases for "prior notification"[39] to national data protection authorities before the processing of personal data could commence, which was a cornerstone of the DPD but was also heavily criticized for being overly bureaucratic and ineffective.[40] It includes new rules addressing explicitly some

[37] *See* C-131/12, Google v. Agencia Española de Protección de Datos ECLI:EU:C:2014:317 (reaffirming an individual's right to have a data processor delete personal data from its files); C-101/01, Lindqvist v. Åklagarkammaren i Jönköping, 2003 E.C.R. I-12992 (defining personal data, holding that webpages are data files and are covered by data protection legislation when they include personal information); C-317/04 and C-318/0404, Parliament v. Council, 2006 E.C.R. I-4795 (agreement with U.S. government on U.S. access to personal data of commercial airline passengers violates EU law); C-468/10 and C-469/10, Asociación Nacional de Establecimientos Financieros de Crédito v. Administración del Estado, 2011 E.C.R. I-12186 (Member States cannot add new rules relating to the lawfulness of processing personal data beyond what is stated in the Directive); C-614/10, Comm'n v. Austria ECLI:EU:C:2012:631 (national data protection authorities need to be fully independent in accordance with the Directive); C-201/14, Bara v. Președintele Casei Naționale de Asigurări de Sănătat*e*, ECLI:EU:C:2015:638 (personal data processing by Member States' government requires legislative measures); C-362/14, Schrems v. Data Prot. Comm'r ECLI:EU:C:2015:650 (safe harbor agreement with U.S. unlawful).

[38] *See* GDPR, *supra* note 3, at Recital 7 ("The objectives and principles of Directive 95/46/EC remain sound, but it has not prevented fragmentation in the way data protection is implemented across the Union, legal uncertainty and a widespread public perception that there are significant risks for the protection of individuals associated notably with online activity. Differences in the level of protection of the rights and freedoms of individuals, notably to the right to the protection of personal data, with regard to the processing of personal data afforded in the Member States may prevent the free flow of personal data throughout the Union. These differences may therefore constitute an obstacle to the pursuit of economic activities at the level of the Union, distort competition and impede authorities in the discharge of their responsibilities under Union law.").

[39] DPD, *supra* note 6, at arts. 18–21.

[40] *See* GDPR, *supra* note 3, at Recital 70 ("While this obligation produces administrative and financial burdens, it did not in all cases contribute to improving the protection of personal data.").

of the perceived threats of digital lives to individual privacy, such as the controversial "right to be forgotten."[41] As a regulation, it indeed Europeanizes data protection and its enforcement, at least formally, with a strengthened European Data Protection Board,[42] the new ability for national regulators to impose much tougher fines on violators,[43] as well as new compliance mechanisms such as "accountability"[44] and mandatory data breach notifications.[45] By the same token, the regulation also empowers Member States to put in place procedures and safeguards, enabling and constraining data processing, thus effectively delegating back to Member States a significant power to shape the regulatory landscape for the processing of personal data within their jurisdiction.[46]

In the context of Big Data, this hybrid nature of the GDPR is quite visible in all three of the core areas of interest: the question of purpose and repurposing, the issue of permissible collection of data and the problem of data retention.

### *A. Repurposing "purpose"*

The GDPR retains the purpose limitation principle of the directive, but it adds a wrinkle. In general, processing of personal data in the European Union requires a clearly defined purpose at the time of data collection, and data cannot be reused for a very different purpose (one that is "incompatible" with the original purpose, in the words of the GDPR).[47] This will continue to constrain Big Data use of personal data in the European Union.

In principle, the two strategies mentioned above to get around these strict limitations will continue to work in the context of the regulation. But important changes introduced in the regulation alter the shape of these strategies.

The first strategy—to get data subjects to consent to a very broad definition of purpose—has become more difficult to implement as the regulation now explicitly specifies the conditions

[41] *Id.* at art. 17.

[42] *Id.* at arts. 64–72.

[43] *Id.* at art. 79.

[44] *Id.* at arts. 5.2, 33–34. *See* Yann Padova, *What the European Draft Regulation on personal data is going to change for companies*, 4 INT'L DATA PRIVACY L. 39 (2014).

[45] GDPR, *supra* note 3, at art. 31.

[46] *Id.* at arts. 83, 86.

[47] A new Article 6.3a offers an exemption for some repurposing, but it only puts into formal legal text guidance that had already in the context of the DPD been issued by the Article 29 WP on purpose limitation and further processing. *See* Article 29 WP.

for consent.[48] In particular, there is now an explicit requirement that consent obliges those that process personal data to request such consent in "an intelligible and easily accessible form," "clearly distinguishable" from other matters and "using clear and plain language."[49] Furthermore, when judging whether consent was given freely, the regulation now prescribes to take into account whether any services rendered by the processing companies were conditioned upon consent, suggesting that such tie-ins between service and consent are indications of a forced and thus invalid consent.[50] Finally, the GDPR in strong language makes clear that data subjects can withdraw their consent at any time.[51]

Taken together, this will make it harder for data processing companies and organizations both to obtain the broad consent necessary for repurposing from the data subjects, and to limit the cases in which data subjects withdraw their consent upon discovering repurposed processing of their personal data. The spirit of the directive has always been to value clear and explicit consent to a transparent request from those processing personal data, but the wording of the directive does not fully reflect that spirit. Given the popularity of this circumvention strategy by data processing entities in the European Union, it is very likely that EU policy makers deliberately chose to restrict the appeal of this strategy in the GDPR.

In contrast, the second strategy—focused on the exemption of data processing for statistical purposes—remains largely intact. In fact, this strategy seems to have been identified by European lawmakers as enabling Big Data without explicitly abandoning the purpose limitation principle: To that end, the use of data for statistical purposes is explicitly deemed to *not* violate the need to stay with a specific purpose.[52] Moreover, the meaning of "statistical purposes" is not narrowly defined in the regulation,[53] and thus can be construed broadly, covering uses not just for the public interest but by private companies for commercial gain as well.[54] This

[48] GDPR, *supra* note 3, at art. 7.

[49] *See id.* at Recital 32 (clarifying what at the very least consent has to capture, namely "the identity of the controller and the purposes of the processing.").

[50] *Id.* at art. 7.4.

[51] *Id.* at art. 7.3.

[52] *Id.* at art. 5.1(b), Recital 40.

[53] *See also id.* at Recital 126c.

[54] It is worth mentioning that an earlier draft version of the GDPR adopted by the Council of ministers in June 2015 aimed to introduce an additional legal ground for the lawfulness of data processing: "Processing of

permits the exemption for the use of data for statistical purposes to be repurposed for Big Data applications. As the GDPR also enables Member States to limit some rights of individuals in the context of these data uses, the repurposing of the statistical purpose exemption seems not accidental, but quite deliberate.[55]

Clearly, as with the directive, using a statistical analysis to influence decision making directly affecting a particular individual would be outside the meaning of "statistical purposes," and also violate the restrictions on "automated individual decision making, including profiling."[56]

The directive requires Member States to put in place appropriate safeguards for data used for statistical purposes. Given the structure as a regulation, one would expect such safeguards to be detailed in the regulation, or at least at the EU level. This, however, is not the case, as the regulation delegates the power to define the safeguards to be in place for the processing of personal data for statistical purposes back to the Member States.[57]

This is quite remarkable, and will likely result in some two dozen different regulatory frameworks throughout the European Union. It will enable some nations to be more permissive of Big Data, and others to be more restrictive. It will certainly reduce the impact of the regulation as a harmonizing force of data protection

personal data which is necessary for archiving purposes in the public interest, or for historical, statistical or scientific purposes shall be lawful subject also to the conditions and safeguards referred to in Article 83." Council of the European Union, 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), No. 9565/15 of 11 June 2015 (emphasis removed). In doing so, the Council was perhaps trying to enable Big Data processing through a *sui generis* legal grounds that was designed to complete the "presumption of compatibility" for further statistical purposes mentioned above. To some extent, the fact that this specific legal ground has been deleted in the final version of the GDPR can be seen as an evidence of the willingness of the European lawmakers to maintain a more conservative approach towards Big Data.

[55] Pursuant to GDPR, *supra* note 3, at art. 83(2) (an in-depth analysis of the various rights of individuals vis-à-vis entities processing personal data is beyond the scope of this Article, but as this example highlights, the scope and breadth of rights of individuals also tracks the changes in the GDPR that facilitate Big Data).

[56] *Id.* at art. 20; *see also* GDPR, *supra* note 3, at Recital 126c, art. 9(2)(i) (setting an explicit limitation).

[57] *Id.* at Recital 126c (Member States "should within the limits of the Regulation, determine statistical content, control of access, specification for the processing of personal data for statistical purposes and appropriate measures to safeguard the rights and freedoms of the data subject.").

regulation in Europe in the context of Big Data, and it will make life harder for companies and organizations operating not just in one but multiple Member States in Europe. This is especially true for online providers wanting to offer their services throughout the European Union. The EU decision makers apparently saw the innovative contributions of innovative e-commerce and Big Data companies as less important compared with the need to improve privacy protection for European citizens.

On the other hand, this delegated setup may help established national players engage in Big Data analysis. For them, due to their size, the cost of adjusting to different national regimes is less than for startups wanting to do online business across Europe's internal borders. Given their national importance, these national incumbents through effective lobbying can hope to influence national frameworks specifying conditions and safeguards.

The regulation foresees that a technical approach—pseudonymization—will play an important role in this context.[58] Pseudonymization refers to the process of purging from data sets data elements that can directly identify a particular data record.[59] Full name, social security numbers or passport numbers are such identifying data elements. After pseudonymization, data is no longer directly and easily identifiable, but can still be referred back to a specific individual when combined with other data and statistical analysis.

Pseudonymization's rise can be seen as a response to technical advances. Full anonymization—purging of all identifiers so that no linking back to individuals is possible—often requires deleting much of the actual data in favor of tabulated results such as sums and averages. Fully anonymous data, of course, is not subject to data protection laws, as it no longer contains any personal data. But the reverse is true too: all data that can conceivably be identified, including by combining data sets, is subject to strict data protection rules.

Therefore, pseudonymization offers a kind of middle ground between directly identifiable personal data subject to all data

[58] *Id.* at art. 83(1). Pseudonymization also appears to be a method of choice in the context of data protection by design and by default (Article 23), data security (Article 30), and as part of codes of conduct (Article 38).

[59] *Id.* at art. 4(3b) (The GDPR defines it as "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person.").

protection rules, and fully anonymized data. By recognizing the value and importance of pseudonymization, the GDPR acknowledges that even seemingly impersonal data can be personal, but it also accepts that such data is potentially valuable and thus under somewhat looser conditions should and can be reused.[60]

The newly phrased Article 6.3a (which oddly is appended to the article on the lawfulness of processing, rather than the principles of processing to which it actually refers) offers concrete guidance for the reuse of data for novel purposes by laying out the balancing test to be undertaken to assess whether and under what conditions a reuse of personal data for novel purposes is lawful in the absence of consent. It explicitly mentions the importance of technical tools, such as pseudonymization.[61] Recital 125 states that any such pseudonymization must minimize personal data as much as possible,[62] but it is important to note that Article 6.3a, which is less strict in this regard, does not reflect such a firm rule.

In short, the GDPR seems to accept that more frequently data will be reused for novel purposes and offers some pathways to that end. In essence, this may lead to a dedicated legal regime governing processing for statistical purposes that captures a significant part of what Big Data is all about. In a nutshell, provided that (i) personal data was originally lawfully collected, and (ii) that the data use is construed as a "statistical purpose processing," "Big Data-esque" data uses can be implemented because further statistical purposes are deemed *per se* compatible with the initial purpose pursuant to Article 5.b. They do not require a specific legal ground as Recital 40 states that when the further purpose is compatible with the initial one, "no separate legal basis is required other than the one which allowed the collection of data" as long as, pursuant to Article 83, "appropriate safeguards" are in place, especially those enabling pseudonymization. Curiously for a regulation aimed at harmonizing data protection, the GDPR delegates to the Member States the extent to which these pathways will be available and useful.

[60] *See* Olivia Angiuli et al., *How to De-Identify Your Data*, 58 COMM. OF THE ACM 48 (2015) (on the process of re-identifying personal data).

[61] GDPR, *supra* note 3, at art. 6.3.

[62] *Id.* at Recital 125 ("These safeguards should ensure that technical and organisational measures are in place in order to ensure, in particular, the principle of data minimisation.").

### *B. Conservative When Collecting*

The GDPR offers formal avenues for the reuse of data, thereby enabling Big Data, but it is much more traditional when it comes to data collection. Collecting personal data, as one element of data processing, continues to require one of the specific reasons enumerated in Article 6, which, with the exception detailed in Article 6.3a mentioned above,[63] mirrors the language of the DPD.

The GDPR delegates the power to lay down specific conditions and safeguards for the processing (including collection) of data again to Member States, but only for purposes of the public interest and processing necessary to fulfil a legal obligation, so the room to maneuver by Member States here is quite circumscribed.

The result is that much like the directive; the GDPR is relatively restrictive when it comes to the collection of data. Here, traditional data protection values such as data minimization, the involvement of the individual, and a strict linkage to purpose remain in place. This may significantly curb the aspirations of Big Data users' aspirations to collect data more comprehensively.

### *C. Reducing Retention Restrictions*

According to traditional data protection principles,[64] data can only be retained as long as it is necessary for the primary purpose for which it is being processed (and thus was collected). Any data retention beyond that requires a new lawful purpose. Due to this preeminence of purpose, consent alone is insufficient to extend the retention of data beyond its initial purpose, as consent to process one's personal data can only be given *in the context of a particular purpose.* The GDPR continues this tradition, when it states that consent can only be given "for one or more specific purposes."[65]

By the same token, the GDPR permits the retention of personal data for longer than absolutely necessary under certain conditions. These conditions are laid out in Article 5(e), stating that longer retention is permitted for, among other reasons, "scientific and historical purposes *or statistical purposes* in accordance with Article 83.1."[66] As arguably most, if not all, of Big Data analysis is statistical in nature, the GDPR offers an explicit pathway for Big Data analyses to work with retained data.

[63] *Supra* Section III.A.

[64] *See* FIPP, *supra* note 17; *see also* OECD Guidelines, *supra* note 17.

[65] GDPR, *supra* note 3, at art. 6.1(a).

[66] *Id.* at art. 5(c) (emphasis added).

It is important, however, to keep in mind that this pathway mandates that any safeguards required by Member States are implemented. This gives members states a somewhat surprising flexibility to choose what regulatory framework Big Data has to comply with, even though the GDPR lays down the key principles and values, including the explicitly mentioned principle of data minimization and the importance of pseudonymity, such a framework has to respect. Data processing entities have to ensure that they comply with the national safeguarding frameworks put in place. This will likely force data processing entities to implement compliance procedures that include a comprehensive assessment component.

To summarize, the key innovation of the GDPR in the retention context is to add a third pathway of data retention. The DPD offered mainly two options: all retention of personal data was limited by tying it to the primary purpose for which it was collected. To escape this straitjacket, personal data had to be completely anonymized, which in practice often meant purging the actual data and only retaining statistical results.

The new third way specified in the GDPR permits retention as long as it is for the purpose of statistical processing and nationally adopted safeguards are implemented, with an explicit emphasis on technical measures such as pseudonymization to reduce the potential harm of prolonged retention.

Taken together, in these three crucial areas, the GDPR is making small, but noteworthy steps towards enabling Big Data in Europe. It is a peculiar kind of Big Data, though, that European policymakers are facilitating: one that emphasizes reuse and permits some retention of personal data, but that at the same time remains very cautious when collecting data. To achieve this result, policymakers were willing to abandon one of the key goals of the GDPR, namely to harmonize and centralize data protection in Europe. In the context of much of Big Data use, the regulatory future in Europe will be shaped by national frameworks, and thus national preferences, values and fears.

## IV. LOOKING BEYOND THE GDPR

If the DPD was Europe's first big, comprehensive statement regarding information privacy and data protection, a kind of Declaration of Independence, the GDPR are like the Articles of Confederation: a step that is far more complex, wordy and ambiguous, full of compromises and lacking in both ambition and coherence. But much like the Articles of Confederation helped generate the fertile soil that prompted the Founding Fathers to design the Constitution not as a logical next step, but as a document breaking with the past and full of simplicity, coherence and vision designed to last, so perhaps the GDPR, too, can be seen as a stepping stone, pointing towards the need to evolve data protection beyond the old paradigm, yet not fully committed to doing so. It also, like the Articles, reflects a time of transition amidst great changes that may have sapped the appetite of overly cautious policymakers to think big(ger).

What could a longer term vision for data protection in the context of Big Data be, especially in Europe? If reuse of personal data becomes common and retention becomes routine; if through the combination of datasets much seemingly innocuous and "impersonal" data can be re-identified; and if individuals continue to exercise their data protection rights, including meaningful consent, as sparingly as they do today by clicking away their rights as an empty formality, it may be necessary to consider replacing the core mechanism of data protection—informational self-determination through the exercise of individual rights.

In its place, one could imagine a mechanism that focuses less on individual consent than on the regulation of permissible and prohibited uses of personal data, protecting individuals irrespective of whether they habitually click the consent button, while also enabling and facilitating accountable and ethical Big Data use.

Such regulation is not new. We already employ it in many domains that have gotten too complex for individuals to comprehend without expert knowledge, and that have important negative externalities. Food, drug and car safety are but three such regulatory fields that immediately come to mind.

While use-based regulation enables processing entities to engage in ethical and accountable uses of personal data without formal consent of the individual, it also saddles them with the explicit duty to act accountably. This requires quite a different approach by data processing entities, shifting away from rituals of consent to deliberate assessment procedures *ex ante*—not just of the benefits but also the potential risks and harms for individuals associated with a particular data use—and the necessity to devise and implement concrete mitigation strategies. The process may be trivial for obvious use cases, but significantly more demanding for complex use scenarios. However, in contrast to individuals, processing entities do have, or at least can obtain, the relevant information to make such assessments. Through the threat of effective enforcement action in case processing entities fail to comply, they will be sufficiently motivated to take the process seriously.

Recently, there has been a groundswell in the literature and among experts on use-based data protection regulation in light of the shortcomings of the existing mechanism, not just in Europe but also in North America.[67] This may be an indication that a part of the expert community is shifting its focus in that direction, mirroring arguments made (unsuccessfully) by some EU policymakers during the GDPR drafting process.[68] This is even more remarkable if we consider that these data protection experts, deeply familiar with existing mechanisms, will have to retool and retrain should a switch towards a use-based mechanism take place.

[67] *See, e.g.*, James Nehf, *Protecting Privacy with 'Heightened' Notice and Choice*, *in* RESEARCH HANDBOOK ON ELECTRONIC COMMERCE LAW (John A. Rothchild ed., forthcoming 2016); Fred H. Cate & Viktor Mayer-Schönberger, *Notice and consent in a world of Big Data*, 3 INT'L DATA PRIVACY L. 67 (2013) (suggesting that notice and choice have become an ineffective mechanism of privacy protection in need to be replaced by a use-based approach). For more critiques on the currently prevailing mechanism of notice and choice *see, e.g.*, Kirsten E. Martin, *Transaction costs, privacy, and trust: The laudable goals and ultimate failure of notice and choice to respect privacy online*, 18 FIRST MONDAY no. 12-2 (2013), http://firstmonday.org/ojs/index.php/fm/article/view/4838/3802; Helen Nissenbaum, *A Contextual Approach to Privacy Online*, 140 DAEDALUS 32 (2011); Robert H. Sloan & Richard Warner, *Beyond Notice and Choice: Privacy, Norms, and Consent*, 14 SUFFOLK U. J. OF HIGH TECH. L. 370 (2014) (agreeing with the critique of notice and choice); Alessandro Mantelero, *The Future of Consumer Data Protection in the E.U. Rethinking the 'Notice and Consent' Paradigm in the New Era of Predictive Analytics*, 30 COMPUTER L. & SECURITY REV. 643 (2014) (suggesting that notice and consent no longer is effective and is in need of at least partial replacement); Joel R. Reidenberg et al., *Privacy Harms and the Effectiveness of the Notice and Choice Framework*, FORDHAM CTR. ON INFO. LAW AND POL'Y (2014), http://moritzlaw.osu.edu/students/groups/is/files/2015/01/Privacy-Harms-and-Notice-and-Choice-01-12-2015-1-4.pdf (mapping areas for which notice and choice are ineffective); CTR. FOR INFO. POL'Y LEADERSHIP AT HUNTON & WILLIAMS LLP, *A Risk-based Approach to Privacy: Improving Effectiveness in Practice* (2014), https://www.hunton.com/files/upload/Post-Paris_Risk_Paper_June_2014.pdf; CTR. FOR INFO. POL'Y LEADERSHIP AT HUNTON & WILLIAMS LLP, *The Role of Risk Management in Data Protection* (2014), https://www.informationpolicycentre.com/files/Uploads/Documents/Centre/Protecting_Privacy_in_World_of_Big_Data_Role_of_Risk_Management.pdf; CTR. FOR INFO. POL'Y LEADERSHIP AT HUNTON & WILLIAMS LLP, *The Role of Risk Management* (2015), https://www.informationpolicycentre.com/files/Uploads/Documents/Centre/Protecting_Privacy_in_World_of_Big_Data_Role_of_Risk_Management.pdf.

[68] *See, e.g.*, Axel Voss & Yann Padova, *We need to make Big Data into an opportunity for Europe*, EURACTIV (June 25, 2015), http://www.euractiv.com/section/digital/opinion/we-need-to-make-big-data-into-an-opportunity-for-europe/.

As we mentioned, the GDPR remains deeply entrenched in the traditional principles of data protection. And yet, when looking closely, one can find numerous signs of a shifting mindset already in place in the regulation. This is quite apparent in the exceptions to the purpose limitation and data minimization principles we detailed above, and the attempt to indirectly define and favor Big Data analyses construed as a "statistical purpose processing."[69] It is also visible in the peculiar structural setup of delegating significant regulatory power back to Member States when it comes to the definition of statistical content, to the conditions and safeguards of Big Data uses of personal data, and to the derogation from data subject's rights.[70] This is also reflected in the repeated references to technical measures, especially pseudonymization, as a kind of third way of enabling some formally prohibited uses of personal data.[71] These are little more than early indicators of change. But they quite possibly point in the direction that data protection legislation is headed, when the GDPR will be replaced with the next evolution of data protection legislation in Europe.

## V. CONCLUSIONS

Information privacy in Europe is experiencing two disruptions. The first, Big Data, is how personal data is being used—shifting where possible from deliberate to incidental collection, from singular use to multiple reuses for varying purposes, and towards longer retention, in order to reap the tremendous latent value in data. The second, the GDPR, is legislative—introducing a new pan-European data protection regulation that aims to evolve Europe's information privacy conception from the 1995 directive.

In this Article we tackled the question of whether these two disruptions amount to a "regime change" for data privacy in Europe. The answer, as is so often the case in the European context, is neither simple nor obvious. On the one hand, the GDPR remains wedded to the core privacy principles of the DPD. On the other hand, there are important changes which directly address some of the key demands of the Big Data community. On the one hand, the GDPR centralizes European data protection legislation; on the other hand, it empowers, to an extent, EU Member States to experiment with regulatory frameworks for data reuse and retention and for the derogation of some of the data

[69] *See* Section III.A.

[70] *See id.*

[71] *See id.*

subject's rights, thus undermining the goal of harmonizing Europe's data privacy framework.

One can describe the situation as muddied, complicated or confusing. One can portray the GDPR as a legislative act lacking clear direction, overall coherence and consistency. One can blame legislative horse-trading that accompanies most major EU initiatives these days. But there is also a more positive view that emerges when looking at legislative acts not as end points of policy debates, but as capturing distinct moments along a historical trajectory. Seen from this vantage point, the GDPR is not breaking with its past, but it is clearly mapping out a pathway into the future that could replace the core mechanism of traditional data protection with a more use-based approach that is much more attuned to a Big Data context. Whether or not Europe is ready to take that next step and change its privacy regime, only time will tell.