PERSONAL DATA EXCHANGES – TOWARDS AN EQUITABLE FOURTH INDUSTRIAL REVOLUTION

*Narayani Anand**

Abstract

This paper aims to examine the effectiveness of the newly adopted Regulation (EU) 2016/679 – popularly known as the General Data Protection Regulation (GDPR) – in protecting data privacy, and analyses the extent to which it prioritises individual interests over those of data aggregators. Three key aspects of data protection, viz. 'notice & consent', 'opting-out' and 'anonymisation & pseudonymisation' have been selected for this analysis. Their presence has then been traced in the GDPR, and compared with the older data protection law in Europe – Directive 95/46/EC, also known as the Data Protection Directive of 1995. Finally, a consumer-centric system of data exchange and management has been proposed vis-à-vis the existing provider-centric model, in the form of a Personal Data Exchange – modelled upon considerations emerging from three separate research approaches – 'Primary Market', 'User Privacy Risk Attitudes' and the 'Personal Information Management System'. This has been proposed as an end towards which the three aspects of

data protection in the GDPR discussed above could be developed.

I. INTRODUCTION

In the age of the internet, what once seemed to be ideas of fiction straight out of Isaac Asimov's works have transformed into reality. The interaction of the internet with common technologies has resulted in outcomes that are altering the way we live. The Executive Chairman of the World Economic Forum (WEF) describes the dawn of this age in words that spell no less than a thrilling anticipation: "we stand on the brink of a technological revolution that will fundamentally alter the way we live, work, and relate to one another. In its scale, scope, and complexity, the transformation will be unlike anything humankind has experienced before."[1]

A transformation "unlike anything humankind has experienced before" will create equivalent challenges. The technical and regulatory frameworks to sustain the Fourth Industrial Revolution are undergoing fundamental changes.

The recently adopted General Data Protection Regulation (GDPR) of the European Union (EU) is being touted as the "world's toughest

*Narayani Anand is a third year law student at Campus Law Centre, Faculty of Law, University of Delhi. The author may be reached at narayani.anand93@gmail.com.

[1]Klaus Schwab, *The Fourth Industrial Revolution: What it Means, How to Respond*, WORLD ECONOMIC FORUM (Jan. 14, 2016), https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.

privacy law".[2] It bolsters the existing provisions for data protection in the EU and is set to harmonize the regulatory framework of its member countries by enacting binding laws.

II. THE EMERGENCE OF BIG DATA

The Fourth Industrial Revolution represents a fundamental change in the way we live, work and relate to one another. It is a new chapter in human development, enabled by extraordinary technology advances commensurate with those of the first, second and third industrial revolutions.[3] We are living on the cusp of opportunity and calamity. The Fourth Industrial Revolution promises technological advancements that can dramatically transform the nature of life on Earth at an unprecedented pace.[4]

This industrial revolution will bring together digital, physical and biological systems. While its conception might still seem abstract, it will be characterised by technologies that will metamorphose the way we live and interact with the physical world. An example of this is the proliferation of artificial intelligence in manufacturing and service delivery.

The key to conceptualizing any of these breakthrough technologies lies in a fascinating concept that is fast taking over the digital world: 'Big Data.' Big Data' is a term that has produced definitional challenges for the sheer variety of contexts it can be understood in. A

[2]David Meyer, *Here Come the World's Toughest Privacy Laws*, FORTUNE TECH (Apr. 14, 2016), http://fortune.com/2016/04/14/eu-parliament-gdpr/.
[3]World Economic Forum, *The Fourth Industrial Revolution*, WORLD ECONOMIC FORUM (June 11, 2018), https://www.weforum.org/focus/fourth-industrial-revolution.
[4]Heerad Sabeti, *The Fourth Sector Is a Chance to Build a New Economic Model for the Benefit of All*, WORLD ECONOMIC FORUM (Sept. 08, 2017), https://www.weforum.org/agenda/2017/09/fourth-sector-chance-to-build-new-economic-model.

definition appearing in a NASA paper, for example, has been argued to be relative and ambiguous[5] for its use of the terms "large" and "more resources" to define, respectively, the size of the data sets and the storage required to fit this data. Further, in a McKinsey study[6] that defines big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze", researchers have acknowledged that "this definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data."

In order to use the most suitable definition for our purpose, it is necessary to emphasize on the regulatory challenges that result from the management of big data. Big Data, therefore, refers to "data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges."[7]

Possibly the first use of the term 'big data' can be traced to the year 1989, when best-selling author Erik Larson penned an article for Harpers Magazine speculating on the origin of the junk mail he received. He wrote that "the keepers of big data say they are doing it for the consumer's benefit. But data have a way of being used for purposes other originally intended."[8] In 1999, the term Big Data

[5]Gil Press, *12 Big Data Definitions: What's Yours?,* WORLD ECONOMIC FORUM (Sept. 3, 2014), https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#1cd6022413ae.

[6]James Manyika et al., *Big data: The Next Frontier for Innovation, Competition, and Productivity,* MCKINSEY & COMPANY: DIGITAL MCKINSEY (May, 2011), https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

[7]The OXFORD ENGLISH DICTIONARY (2013 ed.), http://www.oed.com/view/Entry/18833#eid301162178.

[8] Bernard Marr, *A Brief History of Big Data Everyone Should Read*, WORLD ECONOMIC FORUM (Feb. 25, 2015), https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/.

appeared in research published by the Association for Computing Machinery. One of the aspects that was lamented was the propensity for storing large amounts of data with no way of adequately analysing it. When seen with an undiscerning eye – random sets of data on an individual's social media activity would seem useless. Enterprises, however, are using this data to strike gold, through what is known as data analytics. Data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights, helping organisations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations and higher profits.[9] The difference between enterprises of yesteryears and today is that the latter have understood the importance of capturing all of the data flowing into their businesses and using analytics to extract its maximum value. The Internet of Things (IoT), explained as the concept of "connecting any device with an on and off switch to the Internet and/or to each other",[10] has made it possible to collect and transmit data – in real time. In the past, businesses would collect only a limited type and quantity of data – to be used in making future decisions. This simultaneous collection, transmission and analysis are revolutionizing the way in which enterprises interact with us – the consumers. They now operate faster and stay responsive and are gaining a superior competitive edge.

Consider the case of Aptude,[11] an American IT development firm that uses big data technologies like Hadoop to help its clients harness maximum value through data analytics.

[9]SAS, *Big Data Analytics – What it is and Why it Matters*, SAS INSIGHTS (June 12, 2018), https://www.sas.com/en_us/insights/analytics/big-data-analytics.html.
[10]Jacob Morgan, *A Simple Explanation Of 'The Internet Of Things'*, FORBES (May 13, 2014), https://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#416754321d09.
[11]Aptude, *Big Data Case Study – Hadoop Implementation*, APTUDE (June 12, 2018), https://www.aptude.com/about/case-studies/big-data-case-study-hadoop.

Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.[12]

One of Aptude's clients, a leader in the Transportation and Logistics domain, had their trucks travelling roughly 8 million miles per day. The client needed a method to effectively analyse truck travel patterns to gain an understanding on a myriad of issues including how many "empty miles" were accrued on routes and subsequently make adjustments for more efficient deliveries. Utilising their in-house logistics tracking software, the client had been temporarily storing log files. Due to the massive amount of data being pushed into these files, they were only retaining this data for a short duration. Additionally, since the data was unstructured, developers would have to manually extract, parse, and search the data every time they needed to perform an analysis.

A solution was needed to add structure to these data logs, provide the ability to run ad-hoc queries when issues occurred and perform analytics against the data to improve trucking route efficiency.

After obtaining information through their discovery and requirements gathering process, Aptude architected a big data solution utilising Hadoop in conjunction with a combination of other key open-source components to harness its full potential.

With minimal hardware resources and a collection of open-source software requiring no licensing fees, Aptude realised the Client's big data solution at a fraction of the cost a traditional database solution

[12]SAS, *Hadoop – What is it and Why Does it Matter?*, SAS INSIGHTS (June 12, 2018), https://www.sas.com/en_in/insights/big-data/hadoop.html.

would have required. The Hadoop implementation resulted in cost and time savings, with an additional benefit from the boost in productivity they will achieve with their new analytical assets.

Three key uses of big data analytics to businesses have been identified as:[13]

1. Cost reduction

Cloud based analytics and big data technologies like Hadoop provide notable cost advantages when storing huge amounts of data, as well as in identifying better ways of doing business.

2. Time reduction

In-memory analytics and the processing speeds of Hadoop, along with the ability to analyze new forms of data, enables businesses to analyze information on an immediate basis and make faster decisions.

3. New products and services

Businesses now have the power to tailor their products to fit the customers' needs and preferences One of the most ambitious things an organization can do with big data is to employ it in developing new product and service offerings based on data.

With the multifarious uses of big data- it is evident that its role has expanded significantly.

While in 2013 the IoT market in manufacturing operations was already worth $42.4 billion, it will grow to $98.9 billion by 2018. As with mobile technology 15 to 20 years ago, the IoT revolution is just

[13]Davenport &Dyché, *supra* note 3.

beginning, and over the next two decades it will have a profound impact on businesses, the economy and society.[14]

From the 13 industries that were studied in a research conducted by Tata Consultancy Services(TCS), nearly 79% of the companies used the IoT to track their customers, products, the premises in which they do business with customers, or their supply chains. Perhaps the most significant was the average revenue increase in areas of business where IoT initiatives were deployed – a strong 16% in 2014. In addition, about 9% of firms had an average revenue increase of more than 60%.[15] The CEO of TCS has said that it is because of these developments that he believes data is the new currency.[16]

The value creation offered by big data has become an inevitable asset for companies who want to compete seriously. Research has revealed that a retailer embracing big data has the potential to increase its operating margin by 60 per cent. It also predicts the leveraging of data-driven strategies by, both – established competitors and new entrants – to compete, innovate and capture value.[17]

Data is now part of every sector and function of the global economy and, as an essential factor of production, much of modern economic activity simply could not take place without them.[18]

[14]Natarajan Chandrasekaran, *Is Data the New Currency*?, WORLD ECONOMIC FORUM (Aug. 14, 2015), https://www.weforum.org/agenda/2015/08/is-data-the-new-currency/.
[15]Tata Consultancy Services, *supra* note10. http://sites.tcs.com/internet-of-things/wp-content/uploads/Internet-of-Things-The-Complete-Reimaginative-Force.pdf.
[16]*Id.*
[17]Michael Chui et al., *Big Data's Potential for Businesses*, MCKINSEY & COMPANY (May 13, 2011), https://www.mckinsey.com/mgi/overview/in-the-news/big-data-potential-for-businesses.
[18]*Id*.

III. DATA PROTECTION AND PRIVACY

A. The Need for Data Protection

The data collection activities of businesses have highlighted the pressing need for strong data protection laws. 'Data protection' is defined as the 'legal control over access to and use of data stored in computers.'[19] It is the law designed to protect personal information, which is collected, processed and stored by automated means or intended to be part of a filing system.[20]

Once the data in paper files is converted into a language and format readable by electronic devices, the extraction of personal data from one record and its correlation with the same personal data in another file becomes an easy and inexpensive task. The end-result is a combination that can create a 360 degree online-identity of a person, signalling alarm bells for an individual's privacy.

Consider, for example, the Yahoo! data breach in September 2016. The once dominant Internet giant, while in negotiations to sell itself to Verizon, announced it had been the victim of the biggest data breach in history, likely by "a state-sponsored actor," in 2014. The attack compromised the real names, email addresses, dates of birth and telephone numbers of 500 million users. The company said the "vast majority" of the passwords involved had been hashed using the robust bcrypt algorithm. The breaches knocked an estimated $350 million off Yahoo's sale price.[21]

B. Data Protection In The European Union

[19]THE OXFORD ENGLISH DICTIONARY (2013 ed.) http://en.exforddictionaries.com/definition/data_protection.
[20]Privacy International.
[21]Taylor Armerding, *The 17 Biggest Data Breaches of the 21st Century*, CSO ONLINE (Jan. 26, 2018), https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html.

The strongest and most comprehensive laws are in the countries of the European Union (EU) and European Economic Area that have implemented the 1995 Data Protection Directive. Following the common directive for the region, EU member countries had enacted individual data protection legislations within their national jurisdictions.

After four years of negotiations and formalities, in April 2016, the EU Parliament adopted the "world's toughest privacy law";[22] the General Data Protection Regulation (GDPR). The GDPR will be enforceable from 25 May, 2018, after providing member states with a two-year transition period. Unlike the 1995 Directive that required member countries to pass enabling legislation, the GDPR will be directly applicable and binding on national governments. This will lead to harmonization and better clarity in implementation.

For the purpose of this paper, three aspects of data protection have been briefly examined and their presence has been located in the proposed GDPR. The aspects, viz., 'notice and consent', 'opting out' and 'pseudonymisation and anonymisation' have been chosen for their specific importance to data protection. Their effectiveness as standalone measures in the GDPR has been evaluated.

a) Notice and consent

In the 'Terms of Privacy' laid out by businesses for use of their services, 'notice' implies an informational declaration on the part of the company as to their data collection and processing activities. This may also extend to the notice for third-party data sharing. By clicking 'I agree' on to these privacy agreements, a user, at least theoretically,

[22] *Id.*

consents to the use of their data by the company in the manner so described in their agreement.

The 1995 Directive defined 'consent' in Article 2(h), as "[a]ny freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed." Article 7(a) required that Member States shall provide that personal data may be processed only if the data subject has unambiguously given his consent.

The GDPR has significantly increased the requirements for availing the user's consent, as well as extended to them more rights. Article 7 of the GDPR describes stringent 'conditions for consent' that mandate the controller[23] to be able to demonstrate that the data subject has consented to processing[24] of their personal data. It also requires that the manner for presenting the request for consent be easily distinguishable in an easily understandable form. Further, it provides for the right of the data subject[25] to withdraw such consent, as freely and easily as they give it.

However, aside from this, the GDPR also prescribes the situations in which processing shall be lawful.

Article 6(1) states that processing shall be lawful only if and to the extent that *at least one* of the following conditions apply:

(a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

(b) performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

[23]General Data Protection Regulation (EU) No. 2016/679 of 27 April 2016, art. 4 §§ 33, cl. 7.
[24]*Id*., art. 4 §§ 33, cl. 2.
[25]*Id*. art. 4 §§ 33, cl. 1.

(c) compliance with a legal obligation to which the controller is subject;

(d) for protecting the vital interests of the data subject or of another natural person;

(e) performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; and

(f) for legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Although other provisions in the GDPR are in-built to check the misuse of provisions under Article 6 (such as Recital 32, for example, which provides for the specific acts that constitute consent), the very fact of legalising the processing of personal data in situations besides where such consent is expressly provided, takes away from the primacy of individual consent. This effectively renders the 'consent' clause purely optional for data processing to be lawful, hence negatively impacting individual autonomy. It lends legal backing to the argument most commonly presented by businesses that the consent of users is secondary insofar as data collection and analytics is concerned. This means that organisations can cite "legal obligations" or "contractual performance", for example, and get away with processing a user's data, without their consent. Even with respect to specific conditions such as "legal obligation" under Article 6(1)(c) the recitals make it clear that the relevant "legal obligation" need not be statutory (i.e. common law would be sufficient, if this meets the

"clear and precise" test[26]). A legal obligation could cover several processing operations carried out by the controller so that it may not be necessary to identify a specific legal obligation for each individual processing activity.[27]

b) Opting-Out

'Opting-out' refers to the process of expressly deciding against the collection of information through cookies and sharing of usage and browsing data with third-parties. On websites, pre-ticked boxes that convey the user's consent for information sharing and receiving third-party promotions are the default opt-in options.

Under the 1995 Directive, controllers could rely on "opt-out" and implicit consent in certain situations.[28] The GDPR, however, requires "a statement or a clear affirmative action"[29] by the data subject to signal agreement

Recital 32 of the GDPR states that:

> *"Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement. This could include ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly*

[26]*Id*., Recital 41 §§ 8.
[27]Bird & Bird, *Lawfulness of Processing and Further Processing*, BIRD & BIRD (June 12, 2018), https://www.twobirds.com/~/media/pdfs/gdpr-pdfs/22--guide-to-the-gdpr--lawfulness-of-processing-and-further-processing.pdf?la=en.
[28]Gabe Maldoff, *Top 10 Operational Impacts of the GDPR: Part 3 – Consent*, IAPP (Jan. 12, 2016), https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-3-consent/.
[29]*Supra* note 23, art. 4 §§ 34, cl. 11.

> *indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes or inactivity should not therefore constitute consent."*

Therefore, the Regulation has created additional levels for consent over what was considered legitimate by the 1995 Directive. The latter required consent to be specific to the processing operations and the controller could not request open-ended or blanket consent to cover future processing. Significantly, while consent could be satisfied by an express statement, it also could be inferred from an action or inaction in circumstances where the action or inaction clearly signified consent. Hence, the Directive left open the possibility of "opt-out" consent.[30]

However, through Recital 32, the GDPR removes that possibility by requiring an unambiguous statement implying clear affirmative action on the part of the data subject.

As companies are finding new and improved ways to collect users' personal information and sell it to "third-parties" (most commonly advertisers and marketers), it is becoming increasingly difficult to 'opt-out' of information sharing. The option to limit the sharing of personal information by choosing "opt-out" is not immediately obvious on many websites and applications.

Data as a currency is being traded back and forth by companies to generate millions in profit. Opting out of data brokers and advertising

[30] *Id.*

schemes is notoriously difficult. Other sites make it so you have to provide more information about yourself in order to opt out.[31]

The new law safeguards against this to quite an extent – by mandating a positive "opt-in" mechanism rather than a negative "opt-out" mechanism that would imply consent. This should mean businesses giving special focus to making amply clear the data processing purposes for which consent would be sought.

However, Recital 50 of the GDPR provides for "compatible" operations, citing which consent for subsequent processing operations need not be obtained. These subsequent operations have to be compatible with those for which the data were initially collected. The laws of the EU or Member State may be used to determine and specify the tasks and purposes for which the further processing should be regarded as compatible and lawful.

It also provides certain guidelines that the controller should take into account while determining compatibility, including "any link between those purposes and the purposes of the intended further processing; the reasonable expectations of data subjects based on their relationship with the controller as to their further use; the nature of the personal data; the consequences of the intended further processing for data subjects; and the existence of appropriate safeguards in both the original and intended further processing operations."

While the above guidelines would serve as important safeguards against determining compatibility arbitrarily, Recital 50 provides wide grounds for organisations to manoeuvre outside the limits of consent.

[31]Dave Maass, *How Hard is it to Opt Out of Third Party Data Collection?*, ELECTRONIC FRONTIER FOUNDATION (May 21, 2013), https://www.eff.org/es/mention/how-hard-it-opt-out-third-party-data-collection.

Article 5 contains the principles relating to processing of personal data. Additional processing for reasons of "public interest, statistical purposes, scientific or historical research" will generally be considered compatible under Article 5(1)(b), and, would therefore, be an exception to the requirement for specific consent. Potentially, this exception is quite broad, as – wherever applicable – and read with Article 89 (which contains safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes), even withdrawal of consent by the data subject would not mandate the controller to rectify or erase the data. It would further impact the data subject's right to be notified of and object to processing operations, as well as restrictions on data portability and processing.

c) Anonymisation and Pseudonymisation

The Information Commissioner's Office (ICO) of the UK, an independent regulatory office which reports directly to the Parliament, defines 'anonymisation' as: "the process of turning data into a form which does not identify individuals and where identification is not likely to take place".[32]

Recital 26 of the GDPR defines anonymised data as "data rendered anonymous in such a way that the data subject is not or no longer identifiable." The emphasis in this definition is on stripping the data of any identifiable information in a manner that makes it impossible to get insights on an individual even by the entity that carries out the anonymisation.

[32]Information Commissioner's Office, *Anonymisation: Managing Data Protection Risk Code of Practice, Information Commissioner's Office*, INFORMATION COMMISSIONER'S OFFICE (June 12, 2018), https://ico.org.uk/media/1061/anonymisation-code.pdf.

There is an increasing reliance on anonymisation by organisations in order to broaden the scope of personal data use. Anonymisation of data is carried out to prevent the identification of individuals, organisations and businesses. It addresses ethical concerns regarding protection of people's identities for projects in research as well as for commercial and legal requirements. Common methods include hashing, generating a value or values from a string of text using a mathematical function[33] and encryption the process of using an algorithm to transform information to make it unreadable for unauthorized users.[34]

The Working Party, set up under The Article 29 of the 1995 Directive, had acknowledged that the principles of true data anonymisation were of a very high standard which data controllers often fell short of.

The 1995 Directive, in Rule 26 determining its application, laid down that:

> *"To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".*

The emphasis, thus, was upon identifiability of the data subject from all the means available for likely use by the controller or any other party. If no longer possible, identification would be ruled out and data would thus be considered anonymous while the data protection principles set out in the Directive would no longer apply.

[33]Techopedia,
[34]*Id.*

The GDPR continues this legacy by regarding anonymisation as the highest standard of data protection, thus excluding data that has been anonymised from its purview. Like its predecessor, the Regulation does not apply to anonymised data as defined in Recital 26.

The Regulation brings a novel concept to the data protection law in Europe, by introducing 'pseudonymisation' as a sort of middle-ground aimed at protecting individual privacy while at the same time allowing data controllers to utilise the data.

Article 4(5) of the GDPR defines 'pseudonymisation' as:

> *"the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."*

There is more flexibility in the GDPR vis-à-vis the Directive, in terms of identifiability of individuals. The main point of difference between pseudonymised data and anonymised data is whether there can be re-identification with "reasonable effort".

Even though it falls within the Regulation, some provisions relating to pseudonymised data have been relaxed enough to allow data controllers to benefit from using the technique. Thus, controllers engaging in pseudononymisation of data will find it easier to use it for historical and scientific research purposes as well as in meeting the Regulation's security requirements.

Under the 1995 Directive, the Article 29 Working Party had observed the distinction between the two methods, by stating that "pseudonymisation is not a method of anonymisation" because re-identification remained a possibility, albeit a small one.[35] Therefore, even when the controllers deleted all identifying information on their end, the Directive would apply even if a third-party could reasonably identify the data in future.

In contrast, the GDPR is posed to provide more flexibility, by considering whether re-identification is "reasonably likely".

Pseudonymisation in its present form also facilitates the use and processing of data in excess of its original collection purpose.

Article 6(4) which determines use beyond original purpose for data collected without the data subject's consent, lists "the existence of appropriate safeguards, which may include encryption or pseudonymisation" as one of the factors to be taken into account while determining the compatibility (as discussed under (b.) above). Thus, the GDPR allows controllers who pseudonymise personal data more leeway to process the data for a different purpose than the one for which they were collected.[36]

Further, Article 11 says: "if the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation. This Article also provides that the rights of the data subject contained in Articles 15 – 20, viz. right of access by data

[35]Data Protection Working Party Opinion 05/2014 on Anonymisation Techniques art. 29,

[36]Gabe Maldoff, *Top 10 Operational Impacts of the GDPR: Part 8 – Pseudonymization*, IAPP (Feb. 12, 2016), https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization/.

subject, right to rectification, right to erasure (also known as the 'right to be forgotten') and so on, shall not apply where the controller is able to demonstrate that it is not in a position to identify the data subject. Certain valuable rights of data subjects with regard to control of their data under Articles 15-20 can therefore be waived off simply by the controller demonstrating that he can no longer identify the data subject through the available information.

In any case, the object behind data anonymisation is that the data subject should be nearly impossible to re-identify. The technique, however, falls short of practical and mathematical scrutiny.

It has been shown that 87% of the total population of the United States could be identified by only three markers – their 5-digit zip, gender and date of birth; even when typical data releases contain numerous other fields.[37] In effect, even though these would not be identifiable as standalone data points, storing them together would leave the data subjects susceptible to identification.[38] This, then produces a huge challenge for data controllers seeking to anonymise data.

De-identification – the primary process in anonymisation and pseudonymisation - fails to resist the inferring of sensitive information in boththeory and practice. Attempts to quantify the efficacy of de-intification techniques are unscientific and promote a false sense of security.[39]

In spite of ample scientific evidence to disprove the efficacy of anonymisation and pseudononymisation techniques in data protection,

[37]Sweeney, *supra* note 2.
[38]*Supra*
[39]Narayanan & Felten, *supra* note 1.

the GDPR has wholly excluded anonymised data from its purview, thus leaving millions of people vulnerable to re-identification. This poses an alarming risk to individual privacy, raising serious questions about the rationale behind this move. Further, the GDPR has constructed pseudonymisation regulations with some flexibility – allowing for data controllers to utilise data while also providing for some security measures. The existence of Article 6(4)(f) and Article 11 give great leeway for controllers to process data – a.) for additional purposes without the data subject's consent, and b.) having deleted the identifying information, by simply waiving key rights of the data subjects.

C. Finding A Middle Ground

On examining the efficacy of these three aspects of data protection and their treatment by the GDPR, it is observed that open data is given a preference over data privacy. This is seen, for example, where consent is only one among the six circumstances under which data processing would be deemed lawful[40] , and where – in case of additional processing operations – consent can be altogether done away with, by proving 'compatibility'.[41] Similarly, pseudonymisation has been constructed as a 'middle ground' between security and data use, allowing organisations much elbow-room for harvesting data.

Where "performance of a contract to which the data subject is party"[42] or "legitimate interests pursued by the controller or by a third party"[43]are concerned, data processing would be lawful, whether or not consent of the data subject is obtained. This would facilitate business activities that could involve large-scale mining and harvesting of data – to the extent that appropriate contractual

[40]*Supra* note 23, art. 6 §§ 36, cl. 1.
[41]*Id*. Recital 50 §§ 34.
[42]*Id*., art. 6 §§ 34, cl. 1.
[43]*Id.*

obligations or legitimate concerns pursued by the controller are cited. There is, thus, a tendency to prioritise data use benefits by organisations over data privacy of individuals.

D. Personal Data Exchanges

Any counterbalancing of business interests with those of individuals would be incomplete without a true participation of individuals in the data exchange process. A system of Personal Data Exchanges is proposed to this end. This system would not only streamline the data exchange process with clearly defined data privacy provisions, but would also ensure fair value for both producers as well as users of the data. Whereas traditional data protection models emphasise on protection from a purely control and security perspective, the Personal Data Exchange would deal with data as a commodity, aiming to create and regulate the market conditions necessary for a fair exchange.

d) THE RATIONALE

Data exchange processes and laws have so far placed emphasis on the 'flow', 'storage' and 'use' aspects of data. There is a consequential sidelining of the primary process that is the inception point of all subsequent exchanges – that of data generation. By addressing individuals as 'data subjects', the GDPR fails to address their role as primary producers of data. There is a need to shift the conceptualisation of individuals from subjects to generators and, indeed, owners of their data.

The value harnessed by businesses through big data is a direct outcome of the production of this data by individuals. While traditional business models argue that the existing exchange process

ensures fairness by providing for online services (such as Facebook, for example), in return for the collection and analyses of users' data, it is necessary to consider the actual monetary value in profits harnessed by businesses against the sheer extent and invasiveness of data collection activities.

In a market providing generous returns for effective use of big data analytics by businesses, the individual is the starting point and indeed, an indispensable part of the exchange.

e) THE CONCEPT

It is proposed that true individual participation can only materialise through an independent tech-powered platform – a Personal Data Exchange – that allows individuals to store and control the exchange of their data, thereby enabling them to manage their privacy and optionally monetise parts of their online identity. These would represent the fast-growing economies built on personal data – where businesses share the benefits obtained through user data with its primary generators – the individuals themselves.

f) SOME APPROACHES

i. Creating a Primary Market

Wakenshaw, et al. have argued that a "primary exchange economy" could be created upon internalising these externalities. Such a primary exchange does not yet exist because users do not really exchange personal data; rather giving it away in a dual-step process. Firstly, data is generated through their online actions – which could be, for example, by filling up a form online; and secondly, the automatic transferring away of the data – since the technology used for its collection is created and designed to transfer this data right onto the firm's server. The custodial rights for personal data are therefore held by those collecting information about individuals and not by the

individuals themselves.[44] This data then creates a secondary market between firms, as it is sold for aggregators to gain more insights.

However, it is imperative to appreciate that personal data – generated *by* the individual, *through* technology created by the firm – is co-produced. This co-produced entity could be jointly shared between firm and consumer, if an information-processing platform *owned by* the consumers could store and use their data for their own benefit.

Wakenshaw, et al. propose that an easy, enabling access to such data by both firms and consumers would facilitate a more explicit exchange. This would allow for a wider economy of personal data services – one that would preserve privacy as well as provide value to both, firms and users.

ii.Paying Individuals according to their Privacy Attitudes

In another approach, Aperjis and Huberman have held[45] that there is, in principle, no reason why third parties should not pay individuals for the use of their data. They have then proposed the introduction of a realistic market that would allow these payments to be made while taking into account the privacy attitude of the participants.

It is increasingly accepted that markets 'become' through human effort. It is suggested that "the process of market creation is largely a process of institutionalising certain shared understanding and practices of exchange".[46]

The study focuses on the process of 'legitimation' – lending legitimacy to a new market – through both, cognitive legitimation

[44]*See* Shaprio & Varian, *supra* note 30.
[45]Aperjis& Huberman, *supra* note 1.
[46]Wakenshaw et al., *supra* note 3.

(spread of knowledge of a new venture), and socio-political legitimation (acceptance of a venture by public, government etc., as appropriate given existing norms and laws). The legitimation process would result in the legitimacy of these new products, ideas, practices and institutions.

iii.Personal Information Management Systems (PIMS)

The European Data Protection Supervisor (EDPS)[47] has commented that the prevailing circumstances for processing personal data tend to be unfair to the people whose data is processed. It becomes difficult under the prevailing legal conditions and available technical tools for individuals to exercise their rights, allowing controllers to limit the extent of their liability.

Even where formally having been given some form of a 'notice' and opportunity to 'consent' to general terms and conditions, individuals often find themselves inside a system designed to maximise the monetisation of personal data, which leaves no real choice or control to individuals.[48]

The EDPS, in his Opinion 9/2016, has pushed for Personal Information Management Systems (PIMS). This Opinion explores the concept of technologies and ecosystems aiming at empowering individuals to control the sharing of their personal data. The "vision" of the EDPS as discussed in their Opinion 9/2016 is to create a new reality where individuals manage and control their online identity. It aims to transform the current provider centric system into a human centric system where individuals are protected against unlawful

[47]The European Data Protection Supervisor (EDPS) is an independent supervisory authority whose primary objective is to ensure that European institutions and bodies respect the right to privacy and data protection when they process personal data and develop new policies.
[48]*Supra.*

processing of their data and against intrusive tracking and profiling techniques that aim at circumventing key data protection principles.

It has been argued that providing access rights to customers would be poised to become an inherent service feature delivered to users, instead of being an administrative burden to be complied with.[49] Organisations based on exploiting 'big data' should 'be prepared to share the wealth created by the processing of personal data with those individuals whose data they process'.[50]

This approach, similar to the one propounded by Wakenshaw, et al. puts individuals as holders of their own data. It visualises a 'paradigm shift in personal data management and processing, with social and economic consequences.'

This is contrasted with the existing model of online services where many small providers are owners of a large amount of personal information – thus dominating the market by monetising individuals' personal information as a trade-off for services. The EDPS has correctly recognized the power imbalance that prevails in this circumstance. There is no real concept of choice as the customer has to deal with a 'take it or leave it' set-up. In the presence of a huge 'information asymmetry', there is negligible transparency for users as to what really happens to their personal data.

[49]European Data Protection Supervisor Opinion 7/2015 – Meeting the Challenges of Big Data
[50]*Id.*

The core idea behind the PIMS concept is to transform the current provider centric system into a system centred on individuals able to manage and control their online identity.[51]

At the core of PIMS lies, what the EDPS refers to as 'consent management' – a function that would bring about an automated matching of consumer preferences with requests by providers for personal data. Sufficient detail would be adhered to in expressing privacy preferences after considering a complex collection of possible options. Periodic updating of privacy preferences of customers in this system would ensure that only the most accurate representation of their privacy and risk attitudes is adhered to.

Aperjis and Huberman in their approach have also advocated for differential pricing based on varying risk attitudes – which would enable a fair-pricing mechanism for personal information, for both users and firms. The two approaches are connected in their classification based on privacy preferences and risk attitudes of users. In the process of developing an exchange system – privacy attitudes, therefore, emerge as an important point of consideration.

As a platform incorporated into a model law for the EU, PIMS will ensure compliance with the GDPR for any transfer of personal data beyond the borders of the Union. Creation of similar systems in other jurisdictions will empower users to decide the geographical extent to which they want their data to be shared. It is here that the system will act as a gatekeeper to ensure that the privacy preferences of the user are met. When seen in context of the differential pricing approach, users who allow for a greater geographical net beyond their immediate boundaries for their personal information may be compensated more than others.

[51]*See* Recital 7 GDPR: 'Natural persons should have control of their own personal data'. See also, for example, Doc Searls, The Intention Economy: When Customers Take Charge (Boston: Harvard Business Review Press, 2012).

IV. THE WAY FORWARD

The lacunae in the 'notice', 'consent' and 'pseudonymisation' provisions, as well as others that may emerge upon implementation should be filled with appropriate revisions, which would then serve as a springboard for a consumer-centric approach to the data exchange process. While the implementation of the GDPR is yet to be seen, policymakers must embark on the next steps to chalk out a regulatory framework for Personal Data Exchanges. This will involve – both, market creation and legitimisation – as well as setting fair and appropriate pricing mechanisms.

With promising research emerging in the area of Personal Data Exchanges, it is important that regulatory bodies take into account the next logical step in data protection – ensuring fairness and equity. Personal information should not lose its essence as a user-owned commodity, and its exchange for services should not be seen as an end in itself. In fact, the only means of ensuring that the Fourth Industrial Revolution corrects the imbalances of the earlier ones is through facilitation mechanisms that achieve the three goals of data protection – security, sharing, and monetising – together.

The emerging landscape of PIMS, aiming at putting individuals and consumers back in control of their personal data, deserves consideration, support and further research with a view to contributing to a sustainable and ethical use of big data and to the effective implementation of the principles of the recently adopted GDPR.[52]

[52] *Id.*