EJPLT

European Journal of Privacy Law & Technologies • 2021/1 • ISSN 2704-8012

Bigger is always not better; less is more, sometimes: the concept of data minimization in the context of Big Data.

MD. ABDUL MALEK*
Judicial Officer at Bangladesh Judicial Service

Abstract

With the data landscape of the universe expanding every second every day by leaps and bound, the data value also increases unprecedentedly. Particularly, the disruptive use of data in location tracking, predictive policing, fraud detection, healthcare, advertising media, and entertainment has already revitalized personal data in many ways. But massive amassing of data also gives rise to new issues regarding the Big Data effects, including privacy invasion, data breaches, and cyber threats, etc. Taking effective efforts for mitigating the risks of data explosion thus becomes indispensable for companies, organizations, and societies alike. In such background, this paper attempts to focus on the ways how the data minimization approach mitigates such risks and how this approach as a concept is being incorporated in legal instruments globally. After exploring practical methods of applying data minimization, the paper concludes by delineating the way out of the current dilemmas so created in the face of Big Data.

Keywords: Big Data; Data Minimization; Data Protection Law.

Summary: Introduction. – 1. The Concept of Big Data. – 2. The Concept of Data Minimization. – 3. Data Minimization Concept in Legal Instruments. – 4. Why Data Minimization Approach is Crucial. – 5. Applying the Data Minimization Approach. – 6. Existing Dilemmas and the Way Out. – Conclusions.

Introduction.

The cloud is being fed an enormous amount of data every second. About 2.5 quintillion bytes of data are produced by over 4.1 billion internet users every day. Every day, 95 million photos and videos are shared on Instagram, 306.4 billion emails are sent, and 5 million Tweets are made (Jacquelyn Bulao, 2020). Even, over 350 million photos are uploaded to Facebook each day (Internet World Stats, 2015). By the end of 2020, 44 zettabytes will make up the entire-digital universe. Here is not the end. About 463 exabytes of data will be generated each day by humans as of 2025 (Jacquelyn Bulao, 2020). Right now, although 58.7% of people around the world have access to the internet (ITU, 2015), it will be considerably augmented in the coming days. Hence, it is not unassuming that as our personal data becomes vast, valuable, and sensitive in the world of big data, our privacy concerns and data protection challenges are (and will be) undergoing more intricate situations due to the more advancement of numerous disruptive technologies in the real world.

Basically, the impact of Big Data robustly challenges the data protection regime, and the laws relevant in the field are also fighting for adequately reacting to such deep concerns and challenges. To exemplify them with some questions- if you are not doing hazardous work, should your employer ask for your blood group? When you are in the doctor's chamber for a medical service, should the doctor ask for your religion or ethnicity? Can your recruiter ask for details of health conditions if you are not applying for a manual job? 'In addressing issues akin to these questions in the data ocean, it is the data minimization concept that can help you to stand out. Being a core privacy principle and standard procedure, the concept of data minimization in fact arises as "a solution to the mitigating risks of data Tsunami" (Malek, 2020). This concept is thus being celebrated in the age of Big Data.

As aforesaid, while Big Data promises significant economic and social benefits, it also raises serious privacy concerns and sparks serious debate on its biased, opaque, and discriminating effects. Moreover, data breaches have evidently in recent years grown in number, scope, and magnitude, as it is aptly argued that, "most likely because of the increasing value of digital data about individuals and companies, and the increasing concentration of this data into big-data storage in databases" (Kirtley and Shally-Jensen, 2019). Accordingly,

*Md. Abdul Malek (LLM, Dhaka) works as a career judge at Bangladesh Judicial Service. At the time of publication, the author was working as Judicial Magistrate. He embarked on his professional journey with the Bangladesh Telecommunication Regulatory Commission, the country's national digital space controlling body.

since Big Data has paradoxes: perils alongside its potential, recognizing such paradoxes will help us understand this revolution better (see Richards & King, 2013). Being so is the fact; the problems concerning Big Data management attract some practical approaches or values in the implementation of the data minimization concept.

In such situations, this paper will focus on the imports of the concepts of Big Data and data minimization; and then on how and why the data minimization concept still has the potentials to mitigate the risks. It also highlights the significant jurisdictions of the contemporary globe that has already incorporated the data minimization approach in their legal instruments; and then suggests practical approaches regarding how to benefit from applying the concept in digital space. This paper also concludes by exploring the way out of the current dilemmas that result from data explosion and data minimization.

1. The Concept of "Big Data".

The concept of Big Data is a composite of many concepts (see Brady, 2019). It is rapidly evolving as buzzwords of the moment and one of the most hyped-up terms in recent years. Big Data simply means massive data collection from multiple sources. The term basically "refers to novel ways in which organizations, including government and businesses, combine diverse digital datasets and then use statistics and other data mining techniques to extract from them both hidden information and surprising correlations" (Rubinstein, 2013). In fact, Big Data is a more powerful form of data mining that relies on huge volumes of data, faster computers and also involves, at its core, new analytic techniques to discover hidden and surprising correlations based on artificial intelligence and machine learning to mine the vast amount of at ever-increasing rates; and is being used to inform decisions that affect individuals.

Portraying in a different fashion, Big Data is a methodology rather than a particular selection device. Its' objective is to find 'small patterns' or 'correlations' that reveal new insights or truths. It thus represents a new frontier in how data is processed and used to inform decision-making. It also involves new analytic techniques to discover hidden and surprising correlations based on artificial intelligence and machine learning to mine the vast amount of at ever-increasing rates; it is also being used to inform decisions that have the potentials to affect individuals.

To demystify, it is opportune to posit at the outset that what is 'big' in Big Data is not necessarily the size of the databases, it's the big number of data sources. Big Data is not all about volume; it is big because "the large volume of the real-time data is stored from a variety of sources with random sampling, processed, and produced, in the way of data fusion to create complete-automated-insights" (Steward and Cavazos, 2019). Furthermore, it can be simplified that information becomes Big Data when the volume can no longer be managed with normal database tools. In fact, the concept of Big Data with the power of the machine learning technique is now better known

for its' identifying characteristics 5Vs, i.e., volume (large quantity), veracity (real-time data), velocity (speed), variety (data fusion), and value (worth). Therefore, the right definition of Big Data should in fact be: "Mixed Data" (see Dutcher, 2014; Steward and Cavazos, 2019).

As Kirtley and Shally-Jensen (2019) aptly opined that "the actual or quantifiable measurements of Big Data are still not yet known", it can then be summed up by adding that what is Big Data is hence the 'digital footprint' we intentionally or unintentionally leave behind with every digital step we take. Hence, in the light of the aforesaid dispositions, it can be understood as that it is big, not in terms of its size; instead, it is big because the large volume of the real-time data is stored from a variety of sources with random sampling, processed, and produced, in the way of data fusion to create complete-automated-insights. To clarify further, Big Data is an elaboration of data analysis. Data analysis generates reports on, for example, sales by month. Big Data analysis also examines sales but seeks to find patterns for the effect of time on a day consumers shop, the weather, location of the store, type of credit card, bundle of goods bought, and so on (see Steward and Cavazos, 2019).

Sources of Big Data are all around us and can roughly be divided into business data, human data, and machine data from the internet of things. Big Data examples include credit card transactions, health insurance claims, and online behavior, among others. It now "encompasses a much wider swath of enterprises, and thereby in aid of algorithm improves decision making, enhance efficiency, and, even, increase productivity" (Brynjolfsson, Hitt and Kim, 2011). Indeed, there is evidence that Big Data has led to major breakthroughs in healthcare, more efficient delivery of electrical power, reductions in traffic congestion, and vast improvements in supply chain management (see Tene and Polonetsky, 2013; (Rubinstein, 2013).

Actually, it is the case that "different individuals and organizations access Big Data for different purposes. This, in part, explains why multiple definitions arise in discussions and analyses on the topic" (Kirtley and Shally-Jensen, 2019, p.130). However, while Big Data promises significant economic and social benefits, it also raises serious privacy concerns and sparks serious debate on its biased, opaque, and discriminating effects. Surprisingly, when Big Data is applied in the legal context, as the law has sensitivity, it seriously raises deep concerns and sparks fierce debates for their social, legal, and ethical implications.

2. The Concept of Data Minimization.

Data minimization means the collection and retention of the minimum data possible. It is an "idea that one should only collect and retain that personal data which is necessary" (The International Association of Privacy Professionals, n.d.). The data minimization concept is thus one of the general principles of data protection, which "ideally suggests that the amount of data collected should be the minimal amount of data necessary to conduct businesses" (Kirtley and Shally-Jensen, 2019, p.130). It also refers to measures

performed by organizations to limit the personal data they collect from individuals. In addition to limiting upfront collection, "data minimization also involves deleting or erasing data that is no longer useful as well as setting age limits for data retention" (Dataguise, n.d.).

The concept is basically a concept generally mentioned in the context of protection of personal data, the meaning of which entails the process of gathering solely the data required for fulfilling a particular purpose; that is to say, organizations must only collect the minimum amount of data necessary to accomplish their business purposes; and process personal information that they actually need to achieve the objective of processing the data. It is thus such a principle stating that "data collected and processed should not be held or further used unless this is essential for reasons that were clearly stated in advance to support data privacy" (Experian, n.d.). The concept thus posits the practice of limiting the collection and retention of personal data, and only the minimum amount of data processing is permissible as much as required to carry out the stated purpose (see Tortoise & Hare, 2018) or necessary to accomplish a specified purpose.

The concept also represents best practices with maintaining customer trust and reducing the risk of unauthorized access and other security threats. As a core privacy principle relating to personal data protection law, it envisages that information about an identified or identifiable person can be permissible to process only legitimate processing of personal data. For example, gender may be more relevant than religion or ethnicity if it is for a medical service.

Instead of the 'save everything approach', embracing a data minimization policy as a core principle generally mentioned in the context of protection of personal data is to be prioritized, and unnecessary data is to be discarded, keeping only what is relevant and necessary (see Malek, 2020). Further, that data should be retained only as long as necessary or required by laws or regulations. Organizations or service providers need to ensure that you are not collecting more information than necessary or required. For example, if you only need the e-mail and name of a person to access a service, trying to obtain more information, such as their address or credit card information, is a violation of the General Data Protection Regulation, 2018.

3. Data Minimization Concept in Legal Instruments.

The data minimization concept is notably incorporated in the 2018 General Data Protection Regulation (GDPR), Articles 5, 25, 47, and 89. As one of the seven basic data protection principles under EU data protection law, the very concept lies at the heart of the law embodying as the spirit of the regulatory framework [GDPR, Chapter 2, Article 5 (1) (c)]. According to this regulation, personal data shall be 'adequate, relevant and limited' to what is 'necessary' in relation to the 'purposes' for which they are processed (data minimization) (Art. 5 GDPR, 2018).[1]

[1]Art. 5 (1)c of the GDPR, 2018- "Principles Relating to Processing of Personal Data": adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation') <https://gdpr-text.com/read/article-5/> accessed Summer 1, 2021

Although the term ‘data minimization’ is used multiple times in the GDPR, it does not define “adequate, relevant and limited”, but simply states that “the assessment of what is ‘necessary’ must be done in relation to the purposes for processing” (Torre, 2020). It may also differ from one individual to another, and it thus needs to consider this separately for each individual or for each group of individuals sharing relevant characteristics (see Information Commissioner’s Office, 2019).

Details of the blood groups of the employees from hazardous work are needed in case of an accident. Although adequate safety procedures for preventing accidents are taken, and such data may never be needed, there is much logic that it still needs to hold this information in case of emergency. However, holding by the employer such data of the rest of the workforce is likely to be irrelevant and excessive; that is to say, likely to be unlawful as they do not engage in the same hazardous work. Because the assessment of what data is needed should be based on the purposes of the processing itself, the data controller or processor should never have more data than what it needs to achieve the purposes of the processing (Information Commissioner’s Office, 2019). Then, it means collecting and holding only the minimum amount of personal data needed to fulfill the given purposes.

Another example goes on that ‘in case of gathering the data necessary to answer a particular research question, information that has no value for the research question, should not be collected. For instance, if you would like to build an email subscription list, only collect Name and Email. So, if you are collecting anything more than this (Date of Birth, Religion, etc.), then it might not be compliant with GDPR (Information Commissioner’s Office, 2019)’. Hence, any good data protection practice requires compliance with the principle. As what personal data should be collected and what should not, still remain in discussion, it is completely based on the specific-use-case.

Relevance is another crucial aspect when considering data minimization compliance. Accordingly, apart from a general questionnaire, the recruiters may ask for health conditions that are only relevant to particular manual occupations. An individual applying for an office job should not be asked for such information being irrelevant and excessive to the purpose. For instance, a local hospital to increase patient satisfaction with their care in the pediatric ward collects information at the time of patient check-in; and provides a choice candy to the patient on check out.

Collecting relevant personal information to the stated purpose of providing quality patient care would be considered, but sharing candy preference data with a third-party candy manufacturer would not appear so (see Tortoise & Hare, 2018).

The collection of personal data must not be on the off-chance, i.e., it might be useful in the future, but it may justifiably be so for a foreseeable event (Malek, 2020). Hence, in case of a breach of the data minimization principle, individuals will also have the right to erasure (Information Commissioner’s Office, 2019). Under the GDPR, if the personal data that deems incomplete or insufficient in achieving the purpose of the processing is not ‘adequate’; and considering the context and nature, individuals have the right to complete that data (the right to rectification) (see Torre, 2020). Even further processing

is, however, permissible if the new purpose is not incompatible with the old purpose. Individuals have also the right to get deleted any data that is not necessary for the purpose (the right to erasure, or the right to be forgotten). However, the GDPR also provides for an exception to the data minimization concept, which permits the longer retention of personal data for 'statistical purposes'.

In practice, the data minimization concept forces you to be more conscious about what data you collect. In case of violation of the mandate of the principle, legal actions can be taken by individuals in the European jurisdiction. According to Article 83(5)(a) of the GDPR, the infringements of the basic principles for processing personal data may lead to substantial administrative fines up to €20 million, or 4% of the total worldwide annual turnover. whichever is higher. In France, a fine of 250,000 Euro was imposed on an online retailer 'Spartoo' in August 2020 for a breach of the data minimization principle, among others, full and permanent recording of telephone calls received by customer service employees was held to be excessive. The recording and conservation by the online seller of customer bank details, communicated when orders are placed by telephone, was also 'not necessary' for the intended purpose (see GDPR Enforcement Tracker, 2018).

The data minimization approach is also included in the 2018 California Consumer Privacy Act (CCPA), and the 1988 (Australian) Privacy Act. In fact, "the first regulatory issue concerns personal data protection and consumer protection; then comes to ensuring the application of consumer law on Big Data technologies" (Malek, 2020). It is aptly put that "there are relatively few instances in which data protection authorities have forced technology firms to re-design their software, hardware, or business processes to minimize the processing of data or make it possible for data subjects to use such systems anonymously" (Rubinstein, 2013, p. 74). In conducting a data minimization evaluation, it is necessary to confirm that the collected data is adequate and relevant to the original purposes; and the burden of proving such compliance rests upon the organization.

4. Why Data Minimization Approach is Crucial.

As a risk-management measure, the data minimization approach has become an issue of great importance among information technology stakeholders. It is evident that between the European Union's General Data Protection Regulation (GDPR) and the growing liability of managing large volumes of data in one vulnerable database, businesses are taking a new look at the concept of data minimization (see business.com, 2020). Excess data has a vulnerability for businesses. Hence, it is expedient to point out that to minimize data is to reduce the risk.

Whereas costs are associated with every byte of data store, storing everything either in-house-data centers or cloud archiving is not only unviable but also unnecessary. As a result, traditional forms of data storage shift to lower physical occupancy of data. In fact, holding unnecessary data can bring you more harm than good. For example, as a phone with an overload of apps

and data begins to perform low, a company overflowing with unrequired data in storage begins to stagnate in the long run. Besides, in case of storing too much information, a data breach can be catastrophic when it happens. Even it may lead to charges of criminal negligence. It also costs money and time, and can become dangerous too.

Again, all data are not equally relevant and useful. In essence, much of that data will never be used. Although cloud storage as the latest option for storing data is not expensive, it does not encourage recording all the data and hoarding this excessive data indefinitely that we have. In the existence of such emerging problems, data minimization mitigates both these factors significantly for it stands as a solution to ensuring that we store the data relevant to our purpose. Since more data entails more threats, it is fair to opine that here comes up the relevancy of the notion-the bigger is not always better.

Lack of trust and anonymity has been considered for a long time as the main reason why some always put shopping offline first. By applying the concept of data minimization as a data protection principle, online dealers could reap the great potential benefits of increased online sales (see The International Association of Privacy Professionals, n.d.). The idea 'less data' to be provided may also result in to be much easier, quicker, and user-friendly in an online transaction (see The International Association of Privacy Professionals, n.d.). It should be a benefit for both the company and the individuals. Furthermore, Big Data becomes extremely valuable to the hackers[2] who intent to access to vast amount of data in order to commit fraud or identity theft, or introduce malware to control devices remotely. The risk of data loss and theft is also minimized when only the necessary data is stored up.

In Turkey, a bank faced sanctions for violating the principle of data minimization because the bank provided a six-month account statement of its customer to a civil court when the court only asked for the statement of the last three months (see Malek, 2020). As a result, for harnessing more benefits and mitigating risks, the pursuance of the principle palpably posits that collection or retention of fewer amounts of data is sometimes more in the Big Data world (see GDPR Enforcement Tracker, 2018).

In 2018, under the provision of the GDPR, The Danish Data Protection Authority fined a taxi company 'Taxa', an amount of 1.2 million kroner (US$180,000) for the preservation of the personal data relating to about 9 million individual taxi rides, beyond the lawful two-year retention policy. The regulatory body held that the authorities should also check out whether the data retention policy is duly made and carefully followed (see GDPR Enforcement Tracker, 2018). As it is argued that "companies are concerned about the loss or the leak of data that belongs to them and also about the loss of the personal data with which they have been entrusted" (Kirtley and Shally-Jensen, 2019, p.130), data minimization methods produces significant benefits that in fact include adhering to essential principles of data protection

2 See for example, Betsy Swan, Facial-Recognition Company That Works with Law Enforcement Says Entire Client List Was Stolen, The Daily Beast (February 26, 2020).<https://www.thedailybeast.com/clearview-ai-facial-recognition-company-that-works-with-law-enforcement-says-entire-client-list-was-stolen>.

and EU GDPR compliance; decreasing internal and external threat surface areas; and reducing data storage costs (see Dataguise, n.d.).

Noticeably, privacy policies have thus expanded rapidly on commercial websites as fair practice, including the data minimization approach, to notify users about collecting and using their personal data, even in the absence of any comprehensive law that regulates the substance of privacy policies (see Malek, 2020). Although the data minimization approach is alleged to have crippling effects on Big Data analytics, it basically reduces cost and prevents data breaches from being catastrophic and subject to charges of criminal negligence.

5. Applying the Data Minimization Approach.

Although the practice of the data minimization approach is not an easy thing, it is not difficult too. While collecting personal data, the organizational responsibility comes with 'data classification' that help better to define the data; and decide which personal data is commensurate with the intended purposes or services. For example, in the case of medical service, gender may be more relevant than religion or ethnicity. Consequently, determining what data is absolutely necessary is the first step in a successful data minimization strategy (see business.com, 2020). In the context of European jurisdiction, it is provided that safeguards shall ensure the presence of technical and organizational measures in order to respect the principle of data minimization,[3] in a way that the requirements are not merely directed towards safeguards as such, but towards appropriate ones. Two kinds of safeguards are explicitly mentioned here, namely technical and organizational measures. In this context, practices on the concept of data minimization can be exemplified by pseudonymization and anonymization, i.e., when data is no longer personal data according to the GDPR (Article 2), and thus, falls outside the material scope (see Marcelo Corrales et al. ed., 2017).

The data minimization approach is also considered as DIY (do-it-yourself) data protection practices (see Matzner et al., 2016) that put the responsibility on individuals as the most immediate way to protect their personal data. The term DIY resonates with all measures taken by individual persons to protect their own data by way of, for example, the use of cryptography, pseudonymization and anonymization tools, browser plugins that manage cookies or block tracking and other tools used to minimize data collection.

In the DIY approach, there are passive and active data protection practices. Passive strategies include all strategies relying on withdrawal (opting-out) or data parsimony. Active strategies, on the other hand, encompass the use of privacy-enhancing technologies and taking legal actions. As such, "they serve to build a protected sphere, in which users can perform their selves without

[3]Article 89 (1), GDPR (2018); [Appropriate safeguards concern technical and organizational measures, respecting the data minimization, proportionality, and necessity principle. They may include the possibility of resorting to pseudonymization, where applicable, or further processing "which does not permit or no longer permits the identification of data subjects". The measures involved must take into account the impact on individuals, like physical or psychological potential damages they might suffer]. see GDPR-Text.com. [online] Available at: https://gdpr-text.com/read/article-89/ [Accessed 29 Jan. 2021].

worrying about potential privacy threats" (Matzner et al., 2016). Thus, in furtherance of these measures, the DIY approach requires fostering knowledge and awareness among the data subjects concerning data significance and security.

Moreover, the concept encourages periodically reviewing the data process in order to check whether the stored data is still relevant and adequate for the purposes. If irrelevant or excessive, then delete anything which is no longer needed. In fact, there is a pressing need to have clear and sound data policies on processing, retention, and access by design or by default, and delete and/or archive data on a periodic basis if the organization is holding duplicate and/or unused data. Service providers should have good reasons for asking for specific data. In the case of collecting data that is 'relevant' while creating a personal profile, for example, it is to fix first which data is important.

To exemplify as such, the company should also consider whether it could offer the same feature while collecting less information, such as by collecting zip code rather than precise geolocation. If the company does decide it needs the precise geolocation information, "it should provide a prominent disclosure about its collection and use of this information and obtain consumers' expressed consent. Finally, it should establish reasonable retention limits for the data it does collect" (National Cooperative Freight Research Program, 2019).

As aforesaid, the data minimization approach could boost up online sales in the online retailer's transactions. Hence, there should be no excuse to identify consumers in the name of warranties, potential claims, and taxes. It can be done in the same way as "when buying goods offline, unique product number and transaction details should fully suffice, unless specific laws and regulations require more data for a perfect reason" (The International Association of Privacy Professionals, n.d.). Hence, data classifications, and having clear and sound data policies on processing, retention, and access by design or by default, and reviewing data on a periodic basis can rescue ventures from data explosions.

Strategic data erasure is a core component of the data minimization methodology. User information has a lifespan, which has never been truer than in today's fast-moving digital marketplace. As a result, all data minimization plans should include deletion protocols (see business.com, 2020). User verification and screening through initial assessment procedures in place, organizations may gather only usable information from verified sources (see business.com, 2020).

Hence, there are some pragmatic approaches to minimizing data if the concerned authority respects the principle. On such premise, it is argued that "there are relatively few instances in which data protection authorities have forced technology firms to re-design their software, hardware, or business processes to minimize the processing of data or make it possible for data subjects to use such systems anonymously" (Rubinstein, 2013, p. 74). Considering the imports of the very concept as stated above with pertinent instances and expositions, it can be said that if one asks whether bigger is always better. The answer is in the negative. It is because "even in the realm

of Big Data, companies, and governments are beginning to see the value in a 'less is more' approach" (Marr, 2016).

6. Existing Dilemmas and the Way Out.

There is a dilemma whether data minimization can survive the onslaught of Big Data. In addressing the challenge, Article 23(2) of GDPR creates a more specific obligation for controllers to ensure 'by default' implement mechanisms that data minimization requirements are satisfied. Basically, it is not denying that merely existing of such specific obligation will not suffice because the efficacy of such measure much depends on how it is implemented. So, this new data protection requirement for 'by design and default' may be encouraging, but it still remains in the black box.

There is also another discourse that the "data minimization approach is inimical to the underlying thrust of Big Data, which discovers new correlations by applying sophisticated analytic techniques to massive data collection, and seeks to do so free of any ex-ante restrictions" (Rubinstein, 2013, p.78). It is also claimed that data minimization requirements have crippling effects on Big Data and its associated economic and social benefits. Hence, as it is arguably put that "the regulators should expect to oversee this requirement largely observed in the breach" (see Tene and Polonetsky, 2013). In fact, it is so likely that there should always be a check and balance in the governance mechanism. While companies and organizations are desperate to harness new promises of Big Data analytics, there should be a tool to minimize the associated risk and harm of misuse or abuse of the massive data collection and retention. It is the principle that can operate as such a check on data explosion and encroachment.

It is also argued that "Big Data challenges international privacy laws in several ways as it casts doubt on the distinction between personal and non-personal data, clashes with data minimization, and undermines informed choice" (Rubinstein, 2013). As Big Data is a more powerful version of knowledge discovery in databases or data mining, the dataset, whose size is beyond the ability of the typical database software to capture, store, manage, and analyze, or the nontrivial extractions of implicit, and previously unknown data, potentially poses complex challenges as well ((see Zarsky, 2003; McKinsey Global Institute, 2011).

As aforesaid, providing extra information is not always advantageous for the consumers too. So is also applicable to companies and organizations. In fact, this concept advocates for quality over quantity in a sense, for it imposes limits on the quantity of data that can be processed and requires digressive dada to be discarded.

Even less need may entail less risk of inaccuracy as well. However, discussing in volume the comparative advantages or disadvantages of data minimization and its legal implications requires an interdisciplinary study.

Basically, the GDPR comes as an established, recognized, and rescuing legal instrument which requires companies and organizations to minimize data collection to the extent of relevant, limited, and necessary threshold that can

instead encounter such potential risk of data tsunami. Provably, the data minimization concept is thus still of great relevance in the age of Big Data as it effectively embodies the notion revealing that less is more, sometimes. It is hereby aptly argued that “some information may not need to be collected or shared as initially planned, or the user can be given a choice over which data is processed, based on their functionality needs” (Marcelo Corrales et al. (ed.), 2017).

Conclusions.

The foregoing discourse makes it conspicuous that the preponderance of the data minimization concept largely depends on the consequences which may ensue upon data processing purposes and places. As a result, risk mitigation requires each byte of data processed and stored to be filtered through a series of objectives. If the data does not fit into any of the intended purposes, that should be discarded or deleted. Adopting mechanisms by design, and/or by default, or by cryptography offers the pragmatic way out in the jurisdictions where the concept is incorporated into their legal instruments as a legal principle.

Moreover, building awareness and sensitizing personal data in the Big Data context may also seemingly seem to be a pathway for the minimization initiatives at the social and business domains. At the individual level, the DIY approach is optimally encouraged for minimizing data collection with the tools like browser plugins that manage cookies or block tracking, or pseudonymization and anonymization tools, etc. Even in jurisdictions where regulation or the hard governing law is lacking, the proven benefits of minimizing unnecessary and digressive data collection and storage encourage companies and organizations to be prepared to pursue the approach as a general principle of data protection law. Every democratic society should embrace the same as a virtue as it legitimately limits the task from data collection to data transmission and retention.