GDPR – Challenges for Reconciling Legal Rules with Technical Reality

Mirosław Kutyłowski[1(✉)], Anna Lauks-Dutka[1], and Moti Yung[2,3]

[1] Department of Fundamentals of Computer Science,
Wrocław University of Science and Technology, Wrocław, Poland
{miroslaw.kutylowski,anna.lauks}@pwr.edu.pl
[2] Columbia University, New York, USA
motiyung@gmail.com
[3] Google LLC, New York City, NY, USA

Abstract. The main real impact of the GDPR regulation of the EU should be improving the protection of data concerning physical persons. The sharp GDPR rules have to create a controllable information environment, and to prevent misuse of personal data. The general legal norms of GDPR may, indeed, be regarded as justified and well motivated by the existing threats, however, substantial problems emerge when we attempt to implement GDPR in a real information processing systems setting.

This paper aims at bringing attention to some critical challenges related to the GDPR regulation from this technical implementation perspective. Our goal is to alert the community that due to incompatibility between the legal concepts (as understood by a layman) and the technical state-of-the-art, a literal implementation of the GDPR may, in fact, lead to a decrease in the attainable real security level, thus hurting privacy. Further, this situation may create barriers to information processing environments – including in critical evolving areas which are very important for citizens' security and safety. Demonstrating the problem, we provide a (possibly incomplete) list of concrete major clashes between the legal concepts of GDPR and security technologies. We also discuss possible solutions to these problems (from a technology perspective), and review related activities.

We hope that this work will encourage people to seek improvements and reforms of GDPR based on realistic privacy needs and computing goals, rather than the current situation where people involved in IT projects, merely attempt to only do things that are justified (and perhaps severely restricted) by GDPR.

Keywords: GDPR · Compliance · Privacy · Security

M. Yung—The opinions in this work are personal and do not represent the employers of the authors. The work of the first author has been initiated within the project 2014/15/B/ST6/02837 of Polish National Science Centre.

L. Chen et al. (Eds.): ESORICS 2020, LNCS 12308, pp. 736–755, 2020.
https://doi.org/10.1007/978-3-030-58951-6_36

1 Introduction

Since mid 2018, processing personal data in Europe as well as in some cases also outside Europe has to comply with the General Data Protection Regulation (GDPR) [19] of the European Union. In many other countries (including USA, P.R.C., Japan, Korea, Australia, ...) heavy personal data protection laws have been enforced – however there might be profound differences among these, both in terms of the overall concept and in term of the fine details.

In theory, harsh penalties for not adjusting to GDPR rules on the one hand, and an increased awareness of protection necessity as well as technological advances, on the other hand, should guarantee a swift transition to a world governed by the GDPR rules. However, the emerging reality is quite different: First, the GDPR may be regarded as a *paper tiger* that is not guarding against personal data acquisition (Edward Snowden, 2019), while on the other hand it has created substantial costs for running information systems. Moreover, even in Europe real adjustment to the fundamental principles is far from being the reality. In some areas the situation seems to be alarming (see Sect. 4.3 for some examples: – the situation of AI in Europe and epidemic information processing). Nevertheless, there is almost no discussion about GDPR itself and the necessity for corrections given the experience gathered. As the EU Commission is to present a review on GDPR in 2020, now it is, perhaps, the last moment to express concerns.

Paper Organization. Due to the lack of space we do not provide an introductory guide about the GDPR Regulation – there is an abundant literature on this topic. In Sect. 2 we present challenges for realization of the GPDR rules. These problems are not merely related to ambiguity of a legal text – most of them are due to overlooking the full complexity of issues in the information society. As the devil is in the details, we attempt to provide points that can be described as legal norms that would be appropriate from the technology point of view. In Sect. 3 we present a few general concepts de lege ferenda. In Sect. 4 we report some observed initiatives (or their lack) aimed at adjusting GDPR after two years of experience.

2 GDPR Challenges

2.1 Classification as "Personal Data"

The GDPR regulation says that data falls into the category of *personal data* if and only if it concerns an *identifiable person*. It does not specify the decision context. To make the problem harder, in many practical cases the decision must be automated. Of course, there are evident cases, where the data contain explicit identifiers of data subjects. The hard case is to guarantee that a data piece D does not fall into the category of *personal data*.

A procedure F classifying data as *personal data* or *non-personal data* may take into account the following context as its input:

Option 1: just the data D, by itself,
Option 2: D in the context of all other existing datasets,
Option 3 (hybrid solution): D and a piece of information about the other datasets available to the data controller.

Properties of Option 1: While it is easy to implement, classification as *not-personal data* may turn out to be obviously wrong taking into account the purpose and plausible interpretation of the GDPR. Adopting option 1 would enable easy circumvention of the GDPR principles.

Properties of Option 2: This option is infeasible in practice. Even if access to all other datasets is given, the analysis itself might be extremely complex and infeasible to automate. Definitely, it is very unlikely to get a decision in a real time. Last but not least, in general there is no free access to all other datasets. But even if the access is granted, the permissions are usually granted for a "fair use". The analysis being, in fact, a deanonymization attempt will not necessarily be treated as a "fair use". It can even be regarded as an offense against GDPR.

Properties of Option 3: This option is more realistic for implementation than option 2 (however it is still potentially hard and costly). Unfortunately, it may deliver misleading results from the point of view of GDPR goals.

As we see, every option leads to a kind of deadlock. The main problem is that the attribute *personal data* is regarded as a general property unrelated to the party processing this data. A different approach would be to determine this attribute in the context of the party processing the data. Thereby, the classification of the same dataset D might be different for different parties. Indeed, for instance, if data in D are encrypted homomorphically, then a party P_1 holding D but not the decryption key could perform nontrivial operations on ciphertexts from D without regarding them as *personal data*. At the same time a party P_2 holding the decryption key should regard D as a *personal data* (so even an unauthorized read operation executed by P_2 would be a GDPR violation). An immediate consequence of such an approach is that one cannot permanently classify a given data, the burden to classify data as personal or non-personal is on the party processing the data. (After all: Data is not information on papers which live in a static form forever, data in computations is processed within a context by different parties!)

Rule 1 (*personal data* as an attribute of data and processing party). *A data shall be considered a* personal data *by a party processing it, if and only if this party can identify a physical person related to these data.*

Let us note that Rule 1 has to be used with caution. While for internal processing Rule 1 can be easily applied, it does not automatically enable a data processor P to transfer a non-personal data D to other parties. In this case P has to make sure that D has still the attribute *non-personal* for the receiving party. If this is not the case, then the GDPR rules apply in full for the transfer. These precautions are necessary in particular if P aims to publish D. Then P has

to make sure that by publishing D it will grant an unlawful access to personal data. We discuss these problems in detail in Sect. 2.2.

The proposed approach would have deep consequences for security practice. Assume for instance that party Y has technical capabilities to link the pseudonyms in an anonymous credentials system back to real identities of the users (e.g. thanks to a trapdoor information like the opening authority in group signature schemes which are used for multi-use anonymous credentials, or just thanks to extraordinary computational resources). The system of anonymous credentials can be used as usual, and the regular participants may safely assume the data processed do not fall into the category of personal data. On the other hand, party Y should keep hands off all data from the anonymous credential system, as any operation on these data would mean a violation of the GDPR principles. While Y would have technical possibility to access personal data, they would have no lawful way to take advantage of this knowledge.

In fact, the above situation would be uncomfortable for Y, as it would be necessary to collect a convincing evidence that it has not used its extra knowledge. Indeed, sooner or later it may be revealed that Y had such capabilities and Y might be accused of using it for own advantage. Therefore, for its own interest party Y should create a verifiable system guarding the use of the trapdoor information – e.g. via secret sharing and storing the shares in different locations controlled by different bodies, or an automatic audit of processing activities, and so on. A good (but still academic) example of a scheme that automatically guards against misusing knowledge advantage are fail-stop signatures: even if a powerful party Y can compute discrete logarithms, it cannot use this capability for forging signatures without creating a strong evidence of a forgery.

2.2 Consequences of Processing *Non-personal Data*

If data categorization depends on the processing party, the following problem arises:

Problem 2 (conversion to personal data). It may happen that a party X processes data D, which is *non-personal* from the point of view of X, and creates an output that converts D to *personal data* from the point of view of a party Z.

We propose the following rule:

Rule 3 (impact of processing data). A party X processing *non-personal data* is responsible for all consequences of that processing from the point of view of GDPR.

It can be argued that X can freely process data D, as literally taken these activities are not covered by GDPR. However, indirectly they may have a significant impact of giving access to personal data for other parties:

Example 1. A dataset D with no personal data from the point of view of party A is transmitted to party B in a country where personal data protection rules are not obligatory. This is not directly prohibited by GDPR as the regulation's

scope are exclusively *personal data*. On the other hand, at the same time it may happen that for B or its partner C the data D falls into the category of personal data.

Assume now that B or C de-anonymizes the data D and publishes the resulting data D'. Neither B nor C can be accused of GDPR violations assuming that D' does not concern *offering goods or services* by B or C. Party A also may not be accused directly due to the fact that it has not performed any transfer operation of a *personal data*.

In order to avoid the problems one may adopt the following detailed interpretation of GDPR and the proposed Rule 3:

Rule 4 (Admissibility of data transfer). A may send data D to B iff A can reasonably assume that either:

- D is *not-personal data* for B or any potential partner of B, or
- B complies with the GDPR obligation and has the right to keep this data.

Note that according to Rule 4 it is legitimate to transfer data to party B that complies with the *Privacy Shield* framework.

2.3 Abandoned Data

The case so far ignored by GDPR is the situation when personal data is processed by a party P, but for some reason P becomes inactive (P may become inactive or even cease to exist). The data themselves may be hosted in a system maintained by a third party provider H. What should happen with these data? In an analogous case of a certification authority P, there are detailed rules how to manage data related to the qualified certificates if P terminates its services.

The host H has a dilemma of what to do with the personal data left by P: deleting the data is not allowed, as protection against destroying data is one of the main targets of GDPR. Erasure is allowed only if H is entitled by law or by the data subjects concerned. Keeping the data as they are, is also *processing personal data* in the sense of GDPR and requires appropriate legal grounds.

As long as the data cannot be classified as *personal data* by H, then the situation is slightly less problematic for H – there are no provisions in GDPR that would prohibit erasure of non-personal data. Therefore, in any kind of cloud services it is important for H to apply effective means of pseudonymization. The usual argument in favor of such approach is to guard against misuse of the private data by H. From the point of view of the provider H, the legal safety against supervisory authorities might be regarded as even more important.

Over time it may turn out that the used data protection mechanism becomes ineffective – e.g. due to advances in cryptanalysis. In this case H may be accused of GDPR violations when no countermeasures have been implemented. Note that H may be charged not only when key leakage is its direct fault – GDPR requires to implement appropriate safeguards against third party attacks as well.

2.4 Semantically Neutral Pseudonymization

GDPR attempts to be technologically neutral, however it frequently refers to techniques such as pseudonymization. One may have an impression of pseudonymization being the Holy Grail of personal data protection.

While pseudonymization might be extremely effective and useful in practice, we present an example showing that nontrivial semantic problems may arise. One may hope that pseudonymization or anonymization is a process that, apart from hiding the identity of the data subject, does not change the meaning of the data being processed. This is frequently the case, but not always:

Example 2. Consider in a medical dataset D containing the following note:

Alice suffers the same symptoms as her brother Bob.

There are the following options for pseudonymization of Bob's identity in the medical record of Alice (note that we do not aim to pseudonymize Alice, as she is a patient and her real ID must be available for her physician).

- Replacing *Bob* by a pseudonym X does not change the fact that X can be identified as Bob (provided that Alice has a single brother), so X becomes an *identifiable person* and the data record

$$Alice\ suffers\ the\ same\ symptoms\ as\ her\ brother\ X$$

  still falls into the scope of *personal data* of data subject Bob. The original goal of hiding Bob's identity has not been achieved.
- In order to prevent immediate linking, we replace *her brother Bob* by X:

$$Alice\ suffers\ the\ same\ symptoms\ as\ X$$

  This removes the link to Bob entirely, but semantically this is a different sentence. It merely says Alice is not the only person with these symptoms, while the original data record may indicate possibility of a genetic background.
- One can apply full pseudonymization and store the record

$$Y suffers\ the\ same\ symptoms\ as\ her\ brother\ X.$$

  however this record is useless for the medical treatment of Alice.

Of course, the right for health protection of Alice may overwrite the right of Bob's privacy and this follows directly from GDPR. So, there is an excuse for not applying pseudonymization in medical records. Nevertheless, Example 2 has to show that replacing identifiers by pseudonyms cannot be regarded as a universal tool solving all problems of privacy protection:

Problem 5 (ineffective pseudonymization). There are data records for which pseudonymization either does not hide the identity of a data subject or changes semantic meaning of the data. In both cases pseudonymization fails its purpose.

2.5 Shared Personal Data

GDPR silently assumes that there is a link between a data record and at most one data subject. However, how should the rules apply, if a record D concerns more than one data subject?

Problem 6 (multiple data subjects problem). *Assume that a data record D concerns identifiable data subjects A and B. Then, a consent of which party is required according to GDPR to process D:*

- *a consent of a single party (either A or B), or*
- *a consent of both A and B?*

Example 3. A data record D stored by a controller M concerns data subjects A and B. Then A requests to store D, while B requests to remove D.

What should be the action of M in this situation? If a consent/request of one party suffices to process D, then M gets into troubles:

- if D is erased by M following the request of party B, then what about the right of A for protection of her data from erasure?
- if D is kept by M following the request of A, then what about the right-to-be-forgotten of B?

So whatever M decides to do, it can be accused of violating the obligations formulated in GDPR. If a consent of all data subjects is required, then again M is in a deadlock situation:

- M cannot continue to store D as the consent of all data subjects is missing,
- M cannot erase D as the consent of all data subjects is missing.

Problem 7 (data with multiple data subjects). *It is necessary to agree upon an interpretation or extension of GDPR so that the obligations of the data controller are defined explicitly in case of multiple data subjects of the data.*

An idea for solving this problem might be the following GDPR extension:

Rule 8 (Progressive/regressive data processing). Each data record should have a field or multiple fields *data subject*. The content of this field can be determined manually at the data creation time and can be changed during any editing operation.

The operations on personal data should be classified as progressive and regressive. A progressive operation is an operation that creates a new information contents. A regressive operation is an operation strictly limited to erasing information contents. For progressive operations a consent of all data subjects is necessary (or a corresponding legal reason replacing the consent). For regressive operations a request/withdrawal of the consent by just one data subject is enough to legitimize the operation.

If processing personal data complies with Rule 8, then the following invariant condition holds: if D is stored in a data set then we may assume that all data subjects of D gave their consent to store D in this form.

2.6 Linked Data and Local Categorization

GDPR silently assumes that data D storing personal data of a single party A can be published iff there is a consent of A or a legal reason for this. However, it may happen that there is no legal reason to publish D, A has not given a consent to publish it, but nevertheless the data gets effectively published:

Example 4.
- Alice and Bob gave their consent to publish the following record D: "*Alice and Bob earn together x EUR*" in a public dataset M_1.
- Later Alice gives her consent to insert the record D': "*Alice earns y EUR*" in a public dataset M_2 – a cloud space provided by P_2.

Problem 9 (implicit personal data processing). *In the situation from Example 4, can the provider P_2 of M_2 store D' following the request of Alice?*

There are arguments that P_2 should follow the request as well as deny it:

- The data subject of D' is exclusively Alice, so according to GDPR (and a civil contract with Alice) P_2 may be obliged to store D'.
- On the other hand, publishing D' is equivalent to publication of a record D'': "*Bob earns x–y EUR.*" This happens without a consent of Bob. Moreover, the only data subject for record D'' is Bob.

Problem 10 (semantic analysis). *Is a data processor obliged to perform a semantical analysis of the request concerning data processing having in mind violations of personal data protection of people other than the explicit data subject?*

A positive answer to Problem 10 would raise the question what is the necessary scope of the analysis and how much is the data controller responsible for a misclassification? In practice any analysis may fail. For instance, the access to M_1 might be restricted (e.g. if this is a payed service or a service for a closed set of users). Moreover, P_2 may be even unaware of the existence of the data record D. Even if free access to all datasets that may contain relevant data records D is given, performing necessary analysis may enormously increase the cost and dramatically decrease efficiency of massive data processing. Indeed, an operation on a single data record would require parsing all datasets that may be related to that record (note that the records implying personal information may not be as simple as in the example, but can be, in fact, a complicated analysis of records about subset of people). The result would not be available in real time.

The only solution to this legal deadlock seems to be adoption of the following rule:

Rule 11 (extended context of a consent). *A consent of a party X concerning a data record D in a dataset M should be understood as the right to process D in M regardless of the context that may emerge outside M.*

In Example 4, given Rule 11, the right of Alice to publish D' should be unrestricted. Concurrently, when Bob gives his consent to publish D, he must keep in mind that Alice if free to disclose her personal data.

2.7 Data Aggregation

Assume that party P holds a dataset D containing personal data collected according to GDPR (e.g. the data necessary to run the contracts with the customers of P). Assume that P aggregates data from D: for instance, P may compute the average amount of money spent by the clients of P. Computing such characteristics might be commercially useful for P, but not always are the data processed for the original purpose (realization of a contract or fulfilling legal obligations).

Problem 12 (data aggregation). *Does the result of an aggregation operation fall into the category of personal data? Does aggregation fall into the category of processing personal data – based on the fact that its inputs are personal data?*

Answering the above on the grounds of GDPR is uneasy:

- The aggregated data does not concern a single person, but in the mathematical sense it brings some information about each data subject: e.g. the probability distribution for the salary of Alice might be different than the probability distribution of her salary given the median salary. Note that it may happen that the median belongs to Alice – and the exact amount is revealed. This might be a sensitive information. For instance, the income declared in the previous years is used as an authentication key for a current tax declarations in Poland. It follows that a large anonymity set is not enough to claim that the aggregation result can be revealed without violating cybersecurity conditions (regardless whether we regard it as a violation of GDPR).
- At which moment the aggregated data looses its attribute *personal data*? The legal system only considers a Boolean answer, while in reality there are plenty of possibilities in between.

There might be efforts to find simple shortcuts like: *data concerning a group of more than 10 persons is not personal data*, but it is so easy to misuse such rules in order to bypass the intended personal data protection. The concepts like differential privacy are attractive in theory, but using it in a standard practice could be a nightmare. For instance, what ϵ should be used in the context of ϵ-differential privacy? Or, since differential privacy techniques are based on probabilistic noise, what about cases where the exact sum of private data items has to be calculated for accounting or other commercial purposes?

There is the following legal dilemma:

Problem 13 (data processing classification). *In order to classify a process as processing personal data:*

- *is it enough to find that personal data are used as* ***input*** *in the process? or*
- *the personal data must appear (explicitly or implicitly) in the* ***output*** *of the process?*

A possible pragmatic solution to Problem 13 might be the following:

Rule 14 (narrow definition of data processing). *A data processing P where personal data are included in the input of P shall not be understood as processing of personal data, if the output of P (explicit and implicit) does not contain personal data.*

Rule 14 would be very useful for big data and AI computations like learning models based on private inputs which do not retain private properties. Adopting Rule 14 would also provide a positive answer to the following question:

Problem 15 (right to anonymize). *Is it legal to create a dataset $Anon(D)$ by anonymization of all data records of D?*

A positive answer to Question 15 would solve a lot of problems concerning usage of data that are initially *contaminated* with personal data. This issue has been recognized – for this reason in the current version of GDPR there are exceptions from the general restrictions to process personal data. This touches, in particular, the issue of processing for research purposes (under the provision of respecting the fundamental rights of the data subjects). On the dark side, current exceptions from the general personal data protection rules make room for potential misuses and privacy violations. A rogue party may relatively easily masquerade personal data processing as a research activity. Adopting the right to anonymize data would eliminate the need for the GDPR exceptions in most of the cases occurring in practice. Thereby, one could eliminate the necessity of many exceptions without endangering the research targets.

Let us note that the current GDPR interpretation of European Data Protection Supervisor is not in line with Rule 14 (see Sect. 4.1).

2.8 Quorum Systems

The authors of GDPR seem to have in mind traditional data processing techniques originating from pre-electronic era: it has been silently assumed that each data record is stored in a single physical piece and there is a single party controlling the physical medium used. In such a situation, whenever a data record has to be modified, there is a corresponding physical operation executed on the corresponding physical memory location. GDPR allows a situation of more than one controller (*joint controllers*). In such a case the roles and responsibilities of the controllers have to be strictly defined.

In a quorum system this is not the case. Moreover, a quorum system can be run by a number of independent parties (in fact this is a preferred solution making a user independent from a particular service provider).

In a quorum system a data operation (read, write, erase) is initiated by separate requests to the servers of the system. The key property of the system is that some requests may result in a failure. A quorum system is immune to such failures, as long as the number of failures is limited.

According to GDPR, a data processor has strict obligations concerning data integrity. These obligations are not fulfilled by an individual member of a quorum

system but by the system as a whole. On the other hand, a user has the right to exercise its own rights against any of joint controllers. Thereby, the only safe solution would be to run a quorum system by a single organization. However, this is just what we aim to avoid – a technical dependency on a single organization.

2.9 Secret Sharing

Assume that a personal data D is stored using a secret sharing scheme (e.g. a threshold system k-out-of-n) and that each share is kept by a different party. (This is a favored solution if we have to ensure that no single party can reconstruct the data).

Problem 16. How does the GDPR regulation apply to a party holding a share of a personal data according to a secret sharing scheme?

From the information theoretic point of view the data contained in a share might be purely random (e.g. for schemes based on Lagrange interpolation of polynomials), so one can argue that a share is not a data concerning a data subject. On the other hand, this data collectively with shares from other parties enables reconstruction of the personal data. So which obligations for a data processor apply in this case?

The discussion here is not purely academic – if we adopt the interpretation that a party holding only a share is not processing personal data, then among others the following problems may arise: First, a party keeping a share is no longer obliged by GDPR to protect the share from erasure or modification – indeed, these obligations concern only personal data. Second, after splitting a personal data into shares one could freely store them without a consent of a data subject – as long as each share is stored by a different party. Indeed, no consent is required by GDPR for processing non-personal data. Thereby, it would be very easy to circumvent the personal data protection requirements of GDPR. On the other hand, adopting the interpretation that storing a share falls into the scope of GDPR also creates problems: as long as at least k shares are available in a k-out-of-n threshold system, no complaints are likely. However, in case when less than k shares are intact, then who bears responsibility of the data loss? It is likely that the data subject would sue each share holder that has failed to keep it share. The legal ground would be violation of the GDPR requirement for protecting stored personal data.

2.10 Communication

A standard way of transmitting confidential data over public networks is to send the data encrypted with the key shared by the sender and the receiver (end-to-end encryption). In reality, the communication protocol may involve many operations that are not limited to forwarding the ciphertext to the destination point: the packets may be duplicated, dropped, sent over multiple paths, stored in the spooling systems, additional encryption layers may be imposed, etc. In case

of a standard copyright law, these operations are not regarded as an exploitation of authors' copyright as long as they serve the original purpose of message delivery. Such an exempt does not exist in the case of GDPR: these operations are regarded as processing personal data, as long as the data themselves can be regarded as related to an identifiable person. In almost all communication protocols at least the destination address is explicitly given (the notable exception are the systems like TOR that aim to hide the identity of the communicating parties). In this case the data sent explicitly involves the receiving party. If this is a physical person, then the data fall into the category of *personal data*.

It can be argued that due to encryption the data are unreadable and therefore in some sense "erased." However, even in this case one can argue that encryption is a form of secret sharing (one share is a ciphertext or ciphertexts and one is a key or keys). Consequently, the same concerns apply. Last but not least, there is always some personal data leakage - the communication volume, which reveals the maximal Kolmogorov complexity of the plaintext data transmitted.

Just like in the case of secret sharing, creating a legal framework well capturing existing communication techniques and not creating obstacles for technical improvements is a challenging task. While there are no reports about supervisory authorities putting their hands on existing communication protocols, it cannot be excluded that in the future innovative technologies will be blocked due to such formal reasons. The supervisory authorities may for instance take a position that resilience against traffic analysis is one of GDPR requirements. Anyway, it is relatively likely that provisions of GDPR and the corresponding personal data protection acts can be used as an excuse for protectionist practices in the market of emerging communication networks (5G and beyond).

2.11 P2P Systems

Apart from other problems with distributed systems, there might be problems specific for P2P systems. One of the trouble sources is that the destination server for a given data is not known in advance – it is determined by P2P assignment of logical addresses to the servers. This, in turn, depends on the current state of the system. The owner of the data can be even unaware of the identity of the server storing his data. This prevents using a P2P system by a data controller even for storing encrypted records, as there must be a contract between the data controller and the data processor. For keeping the data by the data subject himself there are similar problems: the party processing the data has, according to GDPR, certain obligations to inform the data subject. However, in many P2P schemes the address of the data subject is not automatically available to the destination server.

Problem 17. *Certain rules of GDPR are hard or infeasible for P2P systems. The problems are caused, among others, by lack of an explicit linking between the party inserting the data and the party storing the data.*

2.12 Blockchain and Append-Only Data Systems

Recently, append-only data systems are gaining popularity, with the Blockchain as perhaps the most prominent example. The most important feature of such systems is that after a data entry is inserted into the system, it cannot be altered or erased. This is fundamental not only for cryptocurrencies, but also for achieving undeniability of transactions in the classical financial systems.

GDPR related problems start, if such a system admits storing personal data. The source of the troubles is the right-to-be-forgotten. No civil agreement may overwrite this right. A party P running the system may defend itself by storing only data that fall into the category of non-personal data. However, this does not guarantee that the problems will be avoided. For instance, a user submitting a signed data s may publish elsewhere a certificate with the public key for signature verification and own real ID. At this moment s becomes personal data, and consequently the user may exercise his right-to-be-forgotten and demand erasure of s. However, this should be technically infeasible and the system provider is trapped to violate GDPR rules.

3 De Lege Ferenda

Apart from solutions suggested in Sect. 2, we present here a few general concepts for privacy protection that would ease deploying pragmatic technical solutions while on the other hand imply high standards of personal data protection.

Processing Personal Data Versus Use of Personal Data. In the current legal situation there are strict rules for processing personal data. On the other hand, the processing itself might be unintentional and/or unconscious. For instance, as noted in Sect. 2.1 it might be non-trivial to determine whether processing of personal data takes place. In the gray area where it is unclear whether the regulation applies, a safe solution is to take precautionary steps and assume that the GDPR rules apply in full. However, such a strategy leads to severe limitations of data processing with profound negative consequences. For instance, it may hinder detection and prevention of financial criminality. Another important field of this kind is controlling the spread of infectious diseases, where adjustment to the rules of GDPR may decrease efficiency of identifying infection chains.

Most negative side effects would be avoided, if the regulation were concentrated on preventing negative consequences of a misconduct. The following limitation involves administrative fines which could serve this goal:

Rule 18. *Administrative fines may be imposed on a party P processing personal data without due diligence, if this results in:*

(a) a profit for P while the rights and freedoms of a data subject are violated, or
(b) a violation of the data subject's rights and freedoms by a third party.

Implementing Rule 18 would prevent any administrative fines when non-compliance with GDPR brings no profit to the data processor while the violations have strictly internal consequences.

The scope of Rule 18 should be understood broadly. It should apply even if the violation addressed by point (b) takes place outside the territorial scope of the GDPR. (Of course, the legal interpretation of our technical desire is in place).

Such an approach would solve some problems, e.g. in the area of data analysis. Even if big data analysis does not fully comply with the GDPR, there will be no sanctions as long as the data processor does not make profit based on a violation. Moreover, the data processor would be able to concentrate on making sure that the rights and freedoms of data subjects are respected, rather than on compliance issues. Last but not least, there might be the cases where the rights and freedoms of data subjects are overridden by other rights and higher values (the common good involving emergencies, is an example).

Responsibility. One of the sources of ineffectiveness of GDPR is focusing on administrative fines payed to the state supervisory authorities. This may result in a situation where the fines secure the state income and not the interests of the potential victims.

An alternative option would be a mechanism of an automatic compensation:

Rule 19. *On demand of a victim, a party getting a profit resulting from a violation of personal data protection rules has to pay a compensation proportional to this profit, independently of who is responsible for the original violation.*

For Rule 19, a default amount could be determined in order to ease risk analysis and eliminate legal disputes. The practice will determine the value of private information based on what a person loss is.

Due to Rule 19, a practical effect on deferring unlawful processing would be more predictable, while the penalties would lose its ad hoc character. In order to avoid penalties, any commercial activity should be supported by procedures that are transparent and provide a self-evident proof of lawful processing. A promising technique in this direction is the SPKI public key infrastructure [7]. What SPKI really provides is a user-centric framework of access control. The SPKI system of delegating access rights would provide a clear and efficient way of determining the right to process personal data.

4 State of Discussion and the Previous Work

While there are many activities regarding the GDPR regulation, there is not much focus on rethinking the general paradigms of GDPR. An implicit assumption is almost always that GDPR is not a subject for discussion and that with some (substantial) efforts one can comply with its requirements. Typically, only narrow aspects of personal data protection in selected IT systems are concerned.

The situation is somewhat surprising taking into account that GDPR itself did not declare itself sacred, and, rather, stipulates regular reviews on GDPR practice by European Commission, taking into account "the positions and findings of the European Parliament, of the Council, and of other relevant bodies or sources". So far, with a few exceptions mentioned below, there are not many such other sources, while the first review is scheduled for 2020. Below we review some ongoing and prior activities; our focus is technical though some related legal issues naturally enter into the discussion.

4.1 Activities of Authorities

European Data Protection Supervisor. Among others, the role of the EDPS is to evaluate regulatory initiatives and provide guidance to the interested parties. So one could expect that the problems of the GDPR implementation are reported in Annual Reports of EDPS. Indeed, there is such a case in the 2019 report [9]: *We received a complaint from a member of staff at an EU institution. He wanted access to his personal data relating to a harassment complaint, submitted against him by a colleague, which had been declared inadmissible.* The part easy to resolve was the invalid ground of data access refusal – protection of a potential victim while the complaint was found inadmissible. A non-trivial issue is that the accusation involved at the same time two other persons in the same institution. Finally, fulfilling the request of the complainant would include the personal data of all accused persons as well as the alleged victim. The position of the EDPS was: *We requested that the EU institution take all reasonable steps to ensure that the complainant's right of access to his personal data was granted.* Literally taken, this opinion indicates that the right to access the data by the complainant overrides the privacy rights of the victim and of the other accused persons. On the other hand, EDPS says that the EU institution should balance the rights and freedoms of the complainant and of the alleged victim. The problem is that such "balancing" is infeasible in a massive processing environment, and creates enormous costs when done manually.

Another interesting issue reported in [9] is *a temporary ban on the production of social media monitoring (SMM) reports* by EASO providing *news on the latest shifts in asylum and migration routes and smuggling offers, as well as an overview of conversations in the social media community relating to key issues.* In this case, the personal data were processed by EASO without a proper legal ground, but the result did not contain personal data, while EASO declared that they took *excess measures to ensure that no personal data was ever stored.* The ban shows that GDPR may block data processing even if the result is not violating privacy and freedoms of data subjects, while there is a clear public interest for processing. So, the current practice does not follow the proposed Rule 14.

EDPS attempts to provide guidance on interpretation of the privacy law. Quite useful are the guidelines proposing a framework for evaluation such aspects as *necessity* and *proportionality* of processing. However, they focus on privacy

protection by the EU institutions and do not cover the problems arising outside the public administration.

ePrivacy. Closely related to GDPR is the ePrivacy regulation proposal [8] aiming at setting barriers for misusing personal data by providers of electronic communication. The general idea of ePrivacy is that a provider of electronic communication cannot use the data of the subscribers except for direct service maintaining. The current process of reaching an agreement upon ePrivacy is slow, with major disputes in the EU.

Quite interesting from the point of view of the problems discussed in Sect. 2.5 is admitting processing user's data by the provider *for the purpose of the provision of an explicitly requested service by an end-user for purely individual use if the requesting end-user has given consent and where such requested processing* ***does not adversely affect fundamental rights and interests of another person concerned*** *and does not exceed the duration necessary* .... This particular condition is a special case of the proposed Rule 8.

US CLOUD Act. Personal data protection is an area of conflict between the EU and the US authorities. The recent US CLOUD Act gives the US law enforcement authorities the right to access personal data processed by US providers overseas. On the other hand, the EU data protection authorities indicate that there is no common understanding regarding law enforcement and the scope of unlawful activities. This creates a very hard situation for some companies – it might be impossible to comply with data access rules of GDPR and CLOUD.

4.2 Academic Research

The academic IT community has devoted a substantial effort to ease implementation of GDPR. Usually, even if the authors point to certain difficulties, some kind of solution within the scope of the current regulation is proposed. Most of the papers present solutions addressed for very specific application areas.

GDPR is a legal concept focusing on essential legal principles to be achieved. Unfortunately, what is observed is that there is a *significant conceptual gap between legal and mathematical thinking around data privacy* (see e.g. [4]).

Compliance. Many researchers provide an evidence that implementation of GDPR creates high organizational costs. Just understanding the requirements by SME is already a substantial problem. Some works provide tools enabling translating the requirements into a form that can be addressed in an algorithmic way [11]. For bigger organizations achieving compliance becomes a very complex task due to its scale. There have been many efforts to ease this process by, say, semi-automated audits (see e.g. [1]).

Privacy Trade-offs. There are examples of substantial problems on the technical side. Among others, the authors point to metadata explosion and conflict with the previous efficiency goals in the area of database design [18], or log size explosion (a read operation has to be followed by a write operation) [16]. In case of technologies like distributed ledger, substantial problems must be solved in order to comply with the right-to-be-forgotten (RtbF) principle [10]. In some cases – like persuasive systems, where a user takes advice by looking at experience of others [17] – the GDPR requirements may block development.

It has also been observed that the procedures introduced due to GDPR may themselves create attack opportunities, despite the original intention. [12] shows such attacks based on abusing the right to access data by the data subject.

New Concepts. A interesting paradigm of the *right-not-to-be-deceived* has been proposed in [14]. It addresses the problem of user manipulation by personalization of the information contents provided to him. This could concern the cases when the user is actually willing to accept the biased information. Enforcing the *right-not-to-be-deceived* would dramatically change the current information processing landscape.

4.3 Industry and Independent Institutions

The controversies about GDPR have started already before it was adopted by the EU. In 2014, the decision by the Court of Justice of the European Union (CJEU) in the Google versus Spain case said that a search engine operator is responsible for processing personal information originating from web pages published by third parties. This decision has been later reflected by the right-to-be-forgotten and the right to data rectification – the most fundamental principles of GDPR. The controversy is that the decision creates an obligation to evaluate data and remove links not only by the parties holding the source data, but also by the parties processing already published data.

Feasibility of the right-to-be-forgotten (RtbF) has been in focus of, both, industry and the academic community [13]. One of the interesting legal concepts was *reputation bankruptcy* – a very controversial one, as the rights and freedoms of a bankrupt person are in conflict with the rights and conflicts of other persons. Balancing these rights or a selective application of RtbF in information systems is a procedural nightmare creating substantial legal risks.

RtbF creates also severe technical problems. It has been pointed out that physical removal of data may take a long time (like 180 days!). Fortunately, the GDPR regulation does not specify a concrete time limit, but refers to an *undue delay*. Nevertheless, the understanding of an *undue delay* may be different for a data controller and a supervision authority. Responding to the demand, some technical concepts have been developed. However, already an overview from 2012 by ENISA [6] has warned about discrepancy between the expectations and the technical reality. Thus far, the technological situation has not changed enough to withdraw this warning.

Another heavily discussed issue was article 22 and the right for a human review of an automated decision. While the regulation could sound reasonable for early IT systems, it ignores complexity of the modern state-of-the-art. For instance, if the decision is based on statistical correlation retrieved from big data, how one can review the outcome and explain the grounds for a data subject [15]?

The AI industry in Europe warns that GDPR creates severe problems for AI computing. Apparently, it has been overlooked by the legislators, as the immediate consequence is that the European industry is losing the competition with USA and China[1]. Article [20] points to the following critical issues:

1. *Requiring companies to manually review significant algorithmic decisions raises the overall cost of AI.*
2. *The right to explanation could reduce AI accuracy.*
3. *The right to erasure could damage AI systems.*
4. *The prohibition on re-purposing data will constrain AI innovation.*
5. *Vague rules could deter companies from using de-identified data.*
6. *The GDPR's complexity will raise the cost of using AI.*
7. *The GDPR increases regulatory risks for firms using AI.*
8. *Data-localization requirements raise AI costs.*

A deep reform of GDPR in the context of AI has been proposed in [2]. The goals proposed are *expanding authorized uses of AI in the public interest, allowing re-purposing of data that poses only minimal risk, not penalizing automated decision-making, permitting basic explanations of automated decisions, and making fines proportional to harm.* The problems for data processing has been also detected by the European authorities (see e.g. [5] for B2B communication problems). The data processing may involve very dynamic manipulation of data, while many GDPR concepts are focused on static data, which is not the emerging reality of modern information systems.

During the corona virus epidemics, it became evident that the current strict protection of personal data and lack of efficacy to make exempts from these rules for the vital public interests, makes GDPR one of the severe obstacles in the fight against corona virus [3]. Facing the catastrophic situation, Italy introduced extraordinary rules for collection and processing on medical data. However, [3] compares such rules to *sticking plaster on a wooden leg* and advocates for an international uniform framework. Definitely, the current situation proves that GDPR has been designed having in mind a standard situation. However, the most important case for evaluating efficacy of a security and safety framework are exceptional and critical situations. For example, while the Google-Apple mechanism for contact tracking carefully attempts to avoid any personal data processing, what about a request of an identifiable person (e.g. a well-known actress) to exercise the RtbF in relation to the BLE signals sent from her smart phone? Obviously, such a request when compared to the common good is a total absurd!

[1] Also in P.R.C. there are opinions pointing to the conflict between the recent cybersecurity law and its data protection chapter on one hand and feasibility of AI data processing.

5 Conclusions

We have presented and exemplified a number of areas and situation which demonstrate shortcomings in the current GDPR rules. Primarily, big data constraints, dynamic and evolving data manipulation and processing, cryptographic tools and techniques, evolving communication and distributed computing configurations, pose new ways by which private data is treated, and pause challenges to GDPR. We also proposed some suggestions (to be considered as initial attempts at remedy), and reviewed current related activities.

References

1. Arfelt, E., Basin, D., Debois, S.: Monitoring the GDPR. In: Sako, K., Schneider, S., Ryan, P.Y.A. (eds.) ESORICS 2019. LNCS, vol. 11735, pp. 681–699. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29959-0_33
2. Castro, D., Chivot, E.: The EU needs to reform the GDPR to remain competitive in the algorithmic economy. Center for Data Innovation (2019). https://www.datainnovation.org/2019/05/the-eu-needs-to-reform-the-gdpr-to-remain-competitive-in-the//-algorithmic-economy/
3. Chivot, E.: COVID-19 crisis shows limits of EU data protection rules and AI readiness. Center for Data Innovation (2020). https://www.datainnovation.org/2020/03/covid-19-crisis-shows-limits-of-eu-data-protection-rules-and//-ai-readiness/
4. Cohen, A., Nissim, K.: Towards formalizing the GDPR's notion of singling out. CoRR abs/1904.06009 (2019). http://arxiv.org/abs/1904.06009
5. Directorate-General for Communications Networks: Study on data sharing between companies in Europe. The European Commission (2018). https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en
6. Druschel, P., Backes, M., Tirtea, R.: The right to be forgotten - between expectations and practice. ENISA (2012). https://www.enisa.europa.eu/publications/the-right-to-be-forgotten/at_download/fullReport
7. Ellison, C.M.: SPKI requirements. RFC **2692**, 1–14 (1999). https://doi.org/10.17487/RFC2692
8. EU Presidency: Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) (amendments) (2020). https://privacyblogfullservice.huntonwilliamsblogs.com/wp-content/uploads/sites/28/2020/02/CONSIL_ST_5979_2020_INIT_EN_TXT.pdf
9. European Data Protection Supervisor: Annual report 2019 (2019). https://edps.europa.eu/sites/edp/files/publication/2020-03-17_annual_report_2020_en.pdf
10. Farshid, S., Reitz, A., Roßbach, P.: Design of a forgetting blockchain: A possible way to accomplish GDPR compatibility. In: Bui, T. (ed.) 52nd Hawaii International Conference on System Sciences, HICSS 2019, Grand Wailea, Maui, Hawaii, USA, 8–11 January 2019, pp. 1–9. ScholarSpace/AIS Electronic Library (AISeL) (2019). http://hdl.handle.net/10125/60145

11. Labadie, C., Legner, C.: Understanding data protection regulations from a data management perspective: a capability-based approach to EU-GDPR. In: Ludwig, T., Pipek, V. (eds.) Human Practice. Digital Ecologies. Our Future. 14. Internationale Tagung Wirtschaftsinformatik (WI 2019), 24–27 February 2019, Siegen, Germany, pp. 1292–1306. University of Siegen, Germany/AISeL (2019). https://aisel.aisnet.org/wi2019/track11/papers/3
12. Martino, M.D., Robyns, P., Weyts, W., Quax, P., Lamotte, W., Andries, K.: Personal information leakage by abusing the GDPR 'right of access'. In: Lipford, H.R. (ed.) Fifteenth Symposium on Usable Privacy and Security, SOUPS 2019, Santa Clara, CA, USA, 11–13 August 2019. USENIX Association (2019). https://www.usenix.org/conference/soups2019/presentation/dimartino
13. Politou, E.A., Alepis, E., Patsakis, C.: Forgetting personal data and revoking consent under the GDPR: challenges and proposed solutions. J. Cybersecur. **4**(1), 1–20 (2018). https://doi.org/10.1093/cybsec/tyy001
14. Reviglio, U.: Towards a right not to be deceived? An interdisciplinary analysis of media personalization in the light of the GDPR. In: Pappas, I.O., Mikalef, P., Dwivedi, Y.K., Jaccheri, L., Krogstie, J., Mäntymäki, M. (eds.) I3E 2019. IAICT, vol. 573, pp. 47–59. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39634-3_5
15. Roig, A.: Safeguards for the right not to be subject to a decision based solely on automated processing (article 22 GDPR). Eur. J. Law Technol. **8**(3) (2017). http://ejlt.org/article/view/570
16. Shah, A., Banakar, V., Shastri, S., Wasserman, M., Chidambaram, V.: Analyzing the impact of GDPR on storage systems. In: Peek, D., Yadgar, G. (eds.) 11th USENIX Workshop on Hot Topics in Storage and File Systems, HotStorage 2019, Renton, WA, USA, 8–9 July 2019. USENIX Association (2019). https://www.usenix.org/conference/hotstorage19/presentation/banakar
17. Shao, X., Oinas-Kukkonen, H.: How does GDPR (General Data Protection Regulation) affect persuasive system design: design requirements and cost implications. In: Oinas-Kukkonen, H., Win, K.T., Karapanos, E., Karppinen, P., Kyza, E. (eds.) PERSUASIVE 2019. LNCS, vol. 11433, pp. 168–173. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17287-9_14
18. Shastri, S., Banakar, V., Wasserman, M., Kumar, A., Chidambaram, V.: Understanding and benchmarking the impact of GDPR on database systems. PVLDB **13**(7), 1064–1077 (2020). http://www.vldb.org/pvldb/vol13/p1064-shastri.pdf
19. The European Parliament and the Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/ec (General Data Protection Regulation). Off. J. Eur. Union **119**(1) (2016)
20. Wallace, N., Castro, D.: The impact of the EU's new data protection regulation on AI. Center for Data Innovation (2018). http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf