Lfd. Nr.

Big Data, Innovation and Privacy

Extended Summary

Community Based Innovation Systems Gmbh (cbased)
SBA Research
Vienna University of Economics and Business Administration

Authors

Clemens Appl (Donau-Universität Krems)
Andreas Ekelhart (sba)
Natascha Fenz (cbased)
Peter Kieseberg (sba)
Hannes Leo (cbased, Koordination)
Sabrina Kirrane (Wirtschaftsuniversität Wien)
Axel Polleres (Wirtschaftsuniversität Wien)
Alfred Taudes (Wirtschaftsuniversität Wien)
Veronika Treitl (Wirtschaftsuniversität Wien)
Christian Singer (BMVIT)
Martin Winner (Wirtschaftsuniversität Wien)

Vienna, December 2017

bmvit

Federal Ministry of Transport, Innovation and Technology

TABLE OF CONTENT

BIG DATA, INNOVATION AND DATA PROTECTION 1

1. Introduction 3

2. GDPR 5

2.1 Explicit consent - ex ante, ex post and beyond 5

2.2 The right to be forgotten 7

2.3 Anonymization - Big Data without restrictions? 8

2.4 Transparency - What exactly happened when? 9

2.5 GDPR compliant development of Big Data applications 10

3. Innovation 11

4. Economic policy recommendations 16

4.1 Initial situation 16

4.2 Measures 17

5. References 23

1. Introduction

The world's most important resource is no longer crude oil, but data, according to the economist's article of 6 May 2017. While this reflects the current assessment of Big Data, the analogy is not really adequate: the volume of data can grow and it is reusable, while crude oil is fundamentally a limited resource. Big Data - a term for which there is no generally accepted definition - is pragmatically seen here as a large amount of data whose analysis requires the use of tools that go beyond standard software applications (e. g. Excel). Collection, storage, analysis, maintenance, search, distribution, transmission, visualization, query, updating and data protection are all challenges due to the size of the database. The aim is to find structures within the data that can explain observed behaviour or predict behaviour[1]. It is not so much about the correlation between data as it is about causal relationships, but also about the algorithms used, the automated decisions based on them and their weak points. Big Data applications are not limited to the economic sphere, but are of great importance in the military, crime, transport, agriculture, health and science sectors. The rules for dealing with Big Data are not uniform.

Big Data receives a lot of attention because

- It is now technically feasible to record, store and analyse the offline and online behaviour of individuals or companies from virtually anywhere in the world via data traces in the network. This includes secret service departments, as well as a number of large US and (potentially also) Chinese companies that have the technical capabilities to do so. However, comprehensive monitoring is already commonplace at a regional level, for example, when retail chains track their customers offline with face/person recognition technology, WLAN and Bluetooth tracking, and connect this data with people via ATM and credit card information. In addition, a large number of companies collect and trade personal data.
- In business and society, transactions often generate sensitive data that requires handling in accordance with data protection regulations. Therefore, adequate technologies and processes are not only required by "data multinationals", but also by all persons and companies dealing with personal data. Applied data protection is therefore not a marginalised challenge but concerns a broad group of actors.
- The digitalisation of all areas of society and the decreasing costs for data acquisition make it much easier and cheaper to record people's behaviour and to digitally map economic, social, administrative and technical processes. New digital products and services, the "Internet of Things" (IoT), offline/online customer tracking in the retail sector, smart homes, smart cities, public administration, digitalised medicine and the digitisation of (traditional) production and distribution processes (keyword: Industry 4.0) generate a massive amount of data and facilitate monitoring and analysis.

The common notion that one can trawl through a pool of data, gradually built up over time, without any technical or methodological restrictions in place, is the source of Big Data fantasies and strategies. Approaches that involve sifting through, restructuring and expanding such existing data, as well as further accumulation of additional data, often form part and parcel of corporate strategies. This approach has to be reviewed with the coming into effect of the General Data Protection Regulation (GDPR 2016/679) in Europe in May 2018. The GDPR is - as much can be stated in advance - a strict,

[1] Definition based on Wikipedia: https://en.wikipedia.org/wiki/Big_data

Europe-wide uniform regime for the handling of personal data. Hence, handlers of personal data face the following challenges and constraints:

- Explicit and informed consent to the processing of data, whereby even storing or anonymising data is considered processing. Practically every manipulation of the data constitutes a further processing, which requires explicit consent of the person whose personal data is affected.
- The data - and this is interesting in connection with Big Data - can be analysed if it is processed anonymously. As already mentioned, consent is required for anonymization.
- GDPR also implies the right to be "forgotten" and thus the deletion of data and information, as well as transparency in data processing and data minimisation - which prohibits retention without valid reason. These demands - however comprehensible and understandable they may be on an abstract level - pose considerable legal/legislative and technical challenges. On the one hand, these provisions are not detailed in depth - partly because technical capabilities are constantly changing and the actual interpretation will arise from court decisions. A problematic circumstance, considering that violations of the GDPR can be punished with a penalty of 4% of the turnover or a maximum of €20 million. On the other hand, there are conflicts of interest between these targets. For example, the demand for transparency and the right to be forgotten may contradict each other. In addition, the technical implementation of many specifications still requires research in order to enable a thorough understanding of the effects.

The GDPR thus establishes stricter rules than, for example, apply in large parts of the USA. As early as during the drafting phase of the legislative text, it was argued that strict data protection hinders innovation. This criticism has not gone away even after the adoption of the GDPR. However, it is known from innovation research that regulations - such as the GDPR - can certainly also contribute to innovations. Strict regulations imposed by environmental legislation have not adversely affected economic growth, but have increased demand for innovative environmental technologies. This creates a market for environmental technologies, and potential producers of these technologies have been encouraged to develop innovative products and services. Can similar things happen through the GDPR?

The question therefore is whether strict data protection hinders companies in innovation or whether it helps them to develop new offerings, business models and technologies that meet both regulatory requirements and the wishes of customers. In principle, it can be assumed that most users attach great importance to data protection, even if they rarely make an effort to ensure that it is enforced.

In order to present evidence to clarify this question, the following methodology was chosen:

1. Legal provisions are examined in light of their technical implications. It will be examined whether technologies and strategies are already in place that help to obtain explicit consent, implement the right to delete data, make data anonymous and allow Big Data to run unrestricted evaluations, while still meeting transparency requirements.
2. The evaluation of technical capabilities is the starting point for assessing the effects on innovation. The main aim is to identify those factors that can have both positive and negative effects. This task necessitates a very pragmatic approach, since the effects of the GDPR can only be observed once it has come into effect in May 2018, and even then, the effects may only become apparent several years later.
3. Recommendations for economic policy are derived on the basis of these analytical steps.

2. GDPR

2.1 Explicit consent - ex ante, ex post and beyond

Obtaining explicit consent to the use or disclosure of personal data from data subjects is a central component of the GDPR. Consent can be obtained electronically, whereby the purpose of use must be clearly indicated. The processing of personal data must not go beyond this purpose. Exceptions exist in relation to processing for archival purposes of public interest, scientific or historical research and for statistical purposes.

The following chapter analyses the technical challenges involved in obtaining explicit consent.

Consent is possible in various ways: 1. general consent as is currently customary with the terms of use of websites, or 2. opt-in, meaning explicit consent to use a particular service.

The provisions of the GDPR pose the following challenges:

- Categorisation: In order not to overwhelm the user or to confront him/her with an unnecessarily lengthy interaction process, it is currently assumed that requests for consent are grouped analogously to the various processing steps, and hence obtained for each individual category.
- Adaptation and revocation: The user must then be in a position - contrary to the status quo, which only allows consent or rejection - to manage their consent in part, and in a fine-grained way. This also includes the possibility to revoke already given consent.
- Comprehensibility: The declarations of consent/requests must be presented in an understandable form and the user must not become overwhelmed. This is particularly challenging in big data scenarios, where complex data processing steps based on non-trivial algorithms and methods have to be explained. Here it is difficult to find the balance between easy explanations, and abstraction. Nevertheless, interested users should be able to understand what a data processing algorithm does with their personal data and how this corresponds to the agreed purpose of use, and also to understand which anonymization procedures and security measures are used to protect their data. Various levels of abstraction and detail are conceivable here.
- Consent after the fact: A further aspect for data processing companies is the inviolability of data for new purposes for which no explicit consent has yet been given. For this reason, it is necessary to create technical methods of obtaining consent (for both the data processor and the users) in an uncomplicated manner.

These declarations of consent are of course different, depending on which processing steps are set and which data is used. The design of a declaration of consent in line with GDPR is therefore not trivial and represents a veritable challenge for many providers.

The precise depiction of the processing steps is of course a challenge when innovations are developed, because it is extremely difficult to indicate ex ante which steps will be taken. From the point of view of data processors, it is desirable that consent to data processing should, as far as possible, also enable the development of new innovative services and business intelligence solutions without having to confront the user with new detailed requests for consent each time - More about this later. It would be desirable on the part of users if the legislature were to stipulate, with regard to service limitations, that, if users do not give or only partially give or withdraw their consent, there remains a guarantee in place that more data protection does not lead to access restrictions that are unfair or discriminatory.

The GDPR makes it clear that informed and specific consent declarations must be obtained. Recent research has raised doubts as to whether this is actually possible. On the one hand, it can be seen that short breaks in the presentation of the data protection declarations/conditions, which are used for the communication of irrelevant information, are sufficient to reduce or even eliminate the attention for or the understanding of data protection declarations/conditions. Users also tend to agree to such terms and conditions without being aware of which of their data is collected, so that their consent is not to be considered as informed consent. On the other hand, it has been observed that an increase in control leads to negligent handling and a higher willingness to take risks. These insights suggest that there is still a real need for research into how to obtain informed consent. At the same time, however, this also means that even with the best of intentions and faultless implementation, the goal of informed consent to the use of personal data is not easy to achieve.

Instead of monolithic, static declarations of consent, there is a need to develop technologies that allow for the dynamic adaptation of consent, with a special focus on legal and also ethical aspects, as well as ease of use and comprehensibility. Such technologies and mechanisms should, on the one hand, enable users to correct data, as well as track and adapt the use of their data[2].

The creation of "best practices" or a European standard for machine-readable consents or processing steps, usage scenarios and categories would be welcome. This would also lead to a wide range of third-party solutions (e.g. apps, bots) that citizens could use to automate or simplify consent procedures.

As a result, many of the machine-readable policy languages described above have already been modelled in RDF (Resource Description Framework), a standard format of the World Wide Web Consortium, for exchanging machine-readable metadata, which can be used to describe and verify usage strategies, legal regulations and business practices, and can be linked to data provenance information and transparency information and events for automatic verification of end-of-life data. There are still gaps, however, as there are different standard RDF schemas/vocabulary for all these aspects, and standards and best practices are lacking for their common use to transparently describe the aspects of the processing of personal data and related policies[3].

In addition, with regard to scalability, there are open research questions about the necessary expressiveness of such policy languages and about the complexity and scalability of corresponding verification methods. Vocabulary describing strategies such as Open Digital Rights Language (ODRL), currently standardised by the W3C's Permissions and Obligations working group, continue to suffer from partial semantic ambiguity that hinders machine processing, or could in practice turn out to be incomplete or insufficient to describe complex strategies for processing personal data. In addition, for secure access to data, RDF must be extended with encryption methods that allow statements to be encoded in a fine-grained manner, which is a largely open research field.

[2] See Kirrane et al. for an overview of Access Control Techniken für RDF.

[3] See also https://www.specialprivacy.eu/images/documents/SPECIAL_D6.3_M9_V1.0.pdf , chapter 4.

2.2 The right to be forgotten

The right to delete data stipulated in the GDPR makes it possible to delete sensitive and personal data from data-processing environments and applications to a certain extent. As simple as this rule may seem, it opens up a challenging and at the same time legally unclear field, because the term "delete" is used differently in data applications. In many data applications and products, this does not mean a final destruction ("physical deletion") of the data, but only an elimination from information processing ("logical deletion"), i. e. the storage space of the data to be deleted is marked as free[4].

However, forensic tools that can recover deleted files have been common for many decades. The success of a recovery depends essentially on whether the blocks have already been reused. This in turn depends on the time elapsed since deletion and the intensity of use in terms of storing new data on the storage medium.

To make this recovery impossible, a set of "final" deletion options has been developed in the past, the most popular of which is randomly overwriting memory areas. There is also an extensive collection of tools for this purpose. Although there are still ongoing discussions about the correct method of overwriting (random, patterns, multiple) – academic laboratories were still able to restore original data, overwritten once using a fixed pattern – this method is nonetheless considered sufficiently effective in practice to assume that the files remain permanently deleted.

However, in extremely critical areas, the physical destruction of data carriers has become commonplace. This is particularly useful if a large amount of data that has been stored for a long time has to be finally destroyed as a whole. However, this access is technically unfeasible for the implementation of the right to be forgotten, which is only concerned with the deletion of individual or a small amount of sensitive information; it is not economically justifiable or technically feasible in highly available systems.

Based on the analyses carried out in this project, the following crucial research questions have emerged. These are not only purely technical questions, but also questions that require an integrated research approach between technical and legal experts:

- Which form of deletion is sufficient and which forensic methods exist? This also includes the development of new forensic methods, which are easy to implement, and which use existing, undeleted meta information for data reconstruction, especially in very complex systems. This is extremely relevant in order to implement the protection of data implied by the GDPR through deletion in a real-world environment. The question is not only limited to "physical" or "logical" deletion, but also covers the handling of backups, security mechanisms, internal (security-) logs, as well as other methods of advanced data management.
- The conflicting objectives with regard to the transparency of data processing must also be clarified. Since it may also be necessary to undo deletions, the deleted content must be kept in appropriate mechanisms. Certain regulations require data not to be deleted so that decisions can be tracked at a certain point in time. The deleted cells are managed in the database in a separate index, the so-called garbage collection, and are therefore analysed not only with

[4] The problem is that although data is no longer logically available - since it can no longer be accessed - it still exists on the storage medium and can therefore be accessed by an attacker. In this case, encryption is an option that cannot be used for legacy systems.

regard to their content, but also the deletion timeline. How to deal with these conflicting goals should be clarified.

In principle, however, the problem of deletion is not limited to data carriers alone, and in many cases cannot be considered purely in the context of the destruction of data carriers. An essential aspect is storage in the cloud or external data centers, where physical storage is not under the control of the data owner. In this case, physical deletion by overwriting is often very difficult to achieve, because the underlying architecture is not known and often only emulated, and hence the respective deletion software cannot function adequately. The same applies to deleting from backups, tape storage, or other mass storage devices, which are not designed to delete individual records and do not always offer this option. In addition, this is also a legal and organisational problem, as backups are often not allowed to be compromised, and are subject to special regulations that must be complied with in order to meet various security certifications. New ways and methods of dealing with such scenarios have to be found on an organisational level.

2.3 Anonymization - Big Data without restrictions?

Anonymization of sensitive data is becoming increasingly important because of the GDPR as it represents an alternative to obtaining explicit consent for the use of personal data. However, it must be noted that anonymization of data is also processing of data, and therefore requires explicit consent.

In order to ensure anonymity, a precise analysis of the information contained in the data is essential in order to be able to unambiguously identify persons from seemingly impersonal information. The data is divided into three types: Identifying data, quasi-identifying data - i.e. data that is unproblematic on its own, but in combination allows identification - and usage data. Anonymization is mainly about the first two groups.

There are a number of strategies and criteria for anonymizing data. The spectrum ranges from synthetic data, cadastres, k-anonymised data and derived methods to differential privacy.

Without going into the various methods at this point, some types of data may lead to a conflict of objectives between strong anonymization and the information content of the data. The greater the anonymization, the lower the information content of the data and thus its use for analytical purposes. However, the usage data - i.e. without personal data - is completely sufficient for many Big Data applications.

One of the main problems with the practical use of anonymization methods is the absence - also with the GDPR - of clearly defined legal requirements for the strength of anonymization (e. g. in the case of k-anonymity, factor k corresponds to the minimum size of the equivalence classes). In addition, data manipulations that ensure anonymity today can be "cracked" as technology advances and would no longer be permissible. Anonymization is therefore a moving target.

In addition, the following questions arise during the practical use of anonymization and have not yet been sufficiently considered:

- The choice of concrete security parameters for anonymization, especially the security factor "k" in the context of k-anonymity or related procedures. The same applies to the use of differential privacy, where the choice of the factor Epsilon has to be clarified. In the case of data perturbation, i.e. the intersection of real data sets with synthetic data, it must be determined which minimum ratio of real data over perturbation data can still ensure the privacy of everyone involved.

- Particularly in the case of sensor data, the classification of sensitive information is not always trivial. Here, it may be necessary to clarify, perchance in industry-specific ways, what characterises quasi identifiers, and establish general criteria for how to recognise and handle them.
- If sensitive data streams arise in the context of internal data processing due to the intersection of (possibly partly sensitive) data, it would have to be clarified when anonymization must be carried out. This is especially important because it is often not possible to merge anonymised data.

Low information content means that the data is much less valuable for Big Data analysis and innovation processes. The missing assignment to data subjects is often not the problem, but the low information content of the data and the resulting low analytical value.

These factors mean that the use of anonymization technologies is associated with a relatively large number of imponderables. In addition, the different approaches require a relatively high level of expertise, which is often not available in small and medium-sized enterprises, and is thus a further limiting factor.

2.4 Transparency - What exactly happened when?

The demand for transparent processing of the data arises directly from the GDPR and thus enables the owner of the data to exercise control over the use of her information. In addition, it can be deduced whether only the agreed data and information were actually used for a data-driven application.

However, this requirement may also conflict with the right to be forgotten, especially if the requirement for transparency is justified by other regulations. Regulations such as SOX (Sarabanes Oxley Act) and Basel II ensure the integrity of the data used in data-driven processing, i.e. they guarantee that the data has not been manipulated at any time. This also applies in particular to external enrichment information, so that it is also possible to re-process data, i.e. it is possible to process data in the way that would have been done at a certain point in time with the information available at that time. This is particularly important in financial transactions, e.g., billing workflows that generate a certain amount of evidence against receivables and disputes.

In enforcing this aspect of the GDPR, it is important to consider which law and what obligations to give priority: The right to be forgotten, or the complete traceability, or even a possible requirement of reprocessing. In our opinion, this will depend on the respective use case and the type of processing.

Transparency in connection with the processing of personal data may also constitute a hurdle, since there are no standard schemes for personal data or generally accepted "best practices" for corresponding granularity of the transparency records. In the field of semantic web research, there are several proposals for ontologies (conceptual schemata, which can be instantiated by means of RDF data) to describe personal data and their provenance (cf. e. g. Bartolini et al. (2015)). However, this and similar works are more academic schemes than standards that could be used directly. It can be assumed that the development and introduction of such standards would contribute to easier workability and verification of transparency records.

A further technical solution is a so-called transparency layer, which is equipped with certain features (completeness, confidentiality, correctness, accuracy, immutability, integrity, interoperability, non-repudiation, correction and deletion, traceability/traceability) and guarantees robust services (high availability and performance, scalability and efficient storage).

This always requires a local transparency layer and a global transparency layer of a third-party organisation that is classified as secure by the data processor and data subject, or a globally managed transparency layer stored in a peer-to-peer architecture.

One possible architecture for a transparency layer is the recently popular blockchain technology (especially through crypto-currencies) to manage and log access to personal data. Blockchain technology is inherently based on peer-to-peer networks and encryption. However, a precise analysis of the non-functional aspects of P2P layers (Ledgers) or block chains as the basis for a transparency layer is urgently needed here: It should be mentioned that voting techniques in the P2P area, which are especially widespread in block chains, allow manipulations when an organization controls more than half of all peers. This is especially important for private blockchains with a small number of peers. Furthermore, since nothing can be deleted in a blockchain per se, it has to be clarified to what extent this technology - for example by using cryptographic deletion (i. e. destruction of the keys) - can simultaneously guarantee transparency and the right of deletion. These questions urgently need answers from science and research, and should be supported via appropriate subsidies.

2.5 GDPR compliant development of Big Data applications

In light of legal requirements, a "naive" development of Big Data applications is not feasible in the context of the GDPR: data collection and data mining analyses are not possible without consent, hence it is not possible to demonstrate the benefits of new features based on the customer's data and an option to withdraw (opt-out).

Based on these findings, however, the following GDPR-compatible procedure is suitable for the development of Big Data applications:

During the development of the data-generating systems, consent to the use of the generated data for subsequent analyses is obtained, stating the processing purpose (e. g. personalised advertising). There are two possibilities: Either the operators of the system obtain prior consent of the data subjects, which must be precisely specified (consent), or they receive consent for anonymization and further processing of the anonymized data. In the latter case, there is no longer any need to specify a precise purpose of use in the analysis and the anonymised data may be used for new analyses unknown at the time of specification.

The advantage of the first option is that arbitrary algorithms can be applied based on the anonymised data. The disadvantage is the loss of information shown in the technical part of the study. From a legal point of view, it would be helpful to have - within the term "anonymization method" - a proper definition of the "anonymization" part, especially since the concrete method used for anonymization will remain subject to state of the art developments.

Since the specific data mining method is not known a priori, it can be assumed that the consent must be adapted more often if the data is not anonymised. Sufficient granularity of the consent within the scope of the application of the GDPR is important. In any case, the comprehensibility for the user must also be taken into account, especially since both anonymization algorithms and data mining algorithms are complex.

In the interest of data minimisation, it is always advisable to separate the operationally processed data, which is only stored for this purpose, from the inputs for Big Data analyses.

When developing applications based on this, it is advisable to employ a group of testers who are willing to submit a declaration of consent tailored to this project, in case consent to data analysis is not sufficient. In this case, it is necessary to obtain appropriate consent declarations from all users once the new application actually goes live.

Compared to the "naive" method of Big Data development, this approach ensures that users are always kept informed about how their data is used. Disadvantages are the limitations on how analysis can be performed, the impossibility to demonstrate to a user the benefits of the new application based on his own data, and the inability to use the "Power of Defaults" through a subsequent opt-out.

However, these disadvantages are counterbalanced by the advantage of greater confidence in the data protection of the offering company, which leads to a higher willingness to agree to the use of data. This is to be seen in particular in connection with the right of deletion and the transparency rules, which make it possible to delete one's own data at any time at a later date. These possibilities also speak in favour of a more generous interpretation of the legal framework when it comes to consent for the anonymization or analysis of data.

3. Innovation

In an increasingly digital society and economy, access to data and the actions it allows are an essential factor in gaining insight into the processes that are taking place. The better it can be analysed, mapped and ultimately forecasted, the greater the value of the data for designing innovations in business, politics, administration, crime prevention, etc.

The different development options for economic activities and product and process innovations resulting from differing data protection regulations are a central question of this study. There is an obvious assumption that data protection regulations influence the development of industries, companies and the public sector (eGovernment) and shape innovation processes. Goldfarb and Tucker (2011) see the greatest effects in the areas of online advertising, eHealth and in-house services.

No specific evidence has been found that strict data protection rules have had a positive impact on innovation. This is of course also due to the fact that the new GDPR is not yet in force. Nevertheless, many observers see Europe's position as a safe haven for personal data, as an opportunity for the digital European economy to develop. At best, the stricter data protection regime in Europe enforces business practices that increase the acceptance of European products and services and thus create competitive advantages. If one considers alleged disadvantages as a challenge, there is room for creative handling of the limitations and for new solutions.

The main focus is on technical solutions that guarantee privacy while still allowing the analysis of data and thus enabling Big Data applications. Priority is given to the development of technologies that only require a small amount of data in order to provide the desired functionality, the possibility of selectively deleting data, and the anonymization of data (see above).

First of all, the question arises as to how GDPR can influence innovation processes. In essence, the following 5 chains of effects have been found, through which data protection can influence innovation activities[5]:

1. Product and service innovation: Innovation processes are search processes in which the environment is interacted with. They are largely open processes. Digitisation has tended to make it easier to bring in relevant expertise - the keyword is open innovation. Innovation processes tend to be more data-driven than in the past. Approaches such as Lean Startup propagate the creation of hypotheses and their data-based validation by potential customers throughout the entire innovation process. It has also been common practice to use market research and internal data sources to gain insight into customer demand patterns and wishes.

In principle, the new General Data Protection Regulation could lead to restrictions here, because some of the data subjects do not give their consent to the use of the data or because a new consent is required for the use of historical data - see below. In practice, there are two cases that can be assessed differently: Startups and established companies.

Start-ups - where the project is the company at the same time - depend on data for their development activities, more precisely, the Lean Startup method propagates a data-driven approach[6]. However, the scope and methods of data collection are rarely called Big Data. Most of these are personal interviews, surveys, web site usage data, etc., which may contain sensitive information, but it is usually also possible to obtain the consent of the data subjects for using data. In the case of start-ups, it is therefore not generally assumed that GDPR hinders innovation processes in the long term.

Established companies have well-functioning business models and thus often have a considerable amount of historical data. Here too, innovation processes are increasingly influenced by lean startup, design thinking[7] and behavioural economics and are thus also strongly driven by interaction with end customers. Here the problem is similar to that of startups.

However, there are a number of options for established companies to deal with the stricter requirements of GDPR. On the one hand, tests can be made with internal users (employees) or a group of test users can be integrated into the development process. It is therefore difficult to argue that data protection fundamentally hinders the development of product and process innovations. This simplistic statement is also not found in any of the scientific articles analysed here. It is therefore obvious that the GDPR requires adaptations in innovation processes, but that these do not in any way lead to fundamental restrictions for innovation activities.

This viewpoint is countered by the fact that innovation processes seldom take place in a well-ordered manner. For example, the principle of data minimisation laid down in the GDPR can work if you know exactly what you want to do, and when, within the innovation process. With this high degree of foresight, it is also possible to clarify the relationship with the data suppliers whose personal data are to be

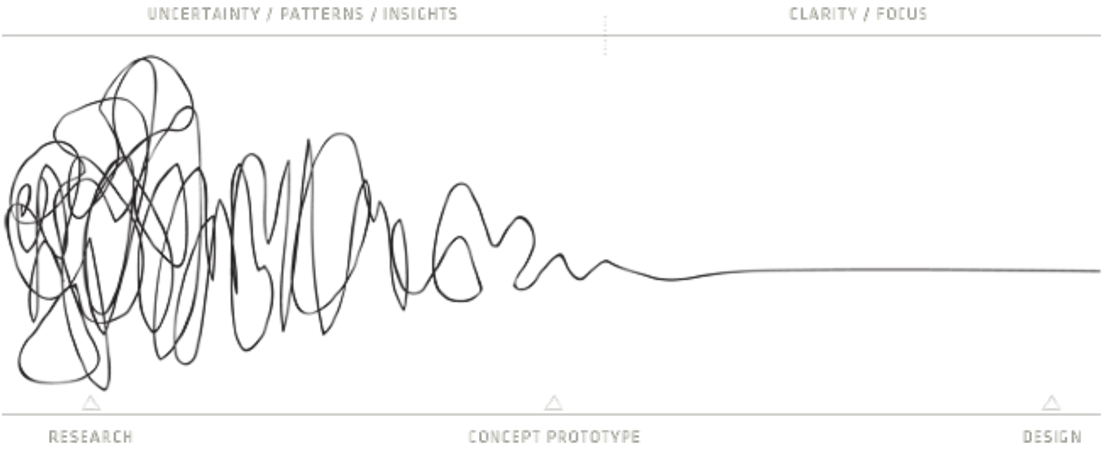
[5] eHealth - the health area - in which strict data protection is a prerequisite for the digital use of data is not dealt here with. See for example Goldfarb - Tucker (2013).

[6] Of course, there are also big data startups - i. e. products and services based on the analysis of large data sets - at the core of their business model. However, these are only a relatively small part of all startups.

[7] According to Wikipedia (https://de.wikipedia.org/wiki/Design_Thinking), Design Thinking tries to find solutions that are convincing from the user's point of view." This means that there is the same goal orientation and also a number of overlaps with Lean Startup in the procedure.

processed. In reality, this situation - when innovation is risky - can be virtually ruled out. Innovation as a process is "messy", characterised by the constant search for combinations that work in the market (see Figure 19), i. e. satisfying a need in a clearly defined market segment and generating enough turnover to allow the development costs and future operating costs to be incorporated. This usually requires several attempts and - in many cases - a pivot, i. e. a fundamental change in the direction of development, different product or service properties, or a different market segment. Particularly when it comes to these fundamental changes, an extensive database would be needed in order to determine the direction. Therefore, innovators strive to collect as much data as possible so that the essential factors for the success of a product can be identified. In this sense - but of course this is not a legally substantiated statement - a "data surplus" can be the most efficient way to develop an innovation. In essence, this illustrates that it is difficult to demand a narrow, ex post interpretation of the principle of data minimisation in innovation processes, because one does not know ex ante which data is needed for the innovation process and therefore there is a tendency to choose a broader search strategy. The same applies, of course, to research and development processes in the academic field. A very restrictive interpretation of the data minimisation principle is not helpful in innovation processes, start-ups and in the scientific field.

Figure 1: Innovation and design processes



Source: Newman

Naturally, one can also solve this problem by obtaining prior consent, if it is formulated broadly enough. The scope and extent of the legal basis for this is currently unclear.

Alternatively, the anonymization of the collected data could avoid restrictions on the use of data. Here, first research results are available on the effects of the right to be forgotten, or on anonymising data on the efficiency of machine learning approaches (Bernd et al. (2016)) - an area that can be described as highly innovative. This mainly concerns the effects these legal provisions have on the underlying algorithms used for data evaluation. It must be noted that especially in the case of self-learning, so-called "intelligent" systems, the data are not only processed but also constitute the system to a considerable extent, i. e. processed and classified data serve as a basis for further analysis, as a so-called "knowledge base". The distortions that occur as a result of modifications to the knowledge base can therefore have crucial ramifications for further processing.

In this example, anonymization has a far greater effect on the usability of data than the selective deletion of important characteristics. However, with regard to the limitations and framework conditions of the results for deletion, it must be pointed out that there is still a great deal more research to be done in this area as well, especially with regard to the anonymization method to be chosen. Nevertheless, this result provides a first indication of the direction in which research is still needed to develop techniques of Privacy Aware Machine Learning (PAML), that can keep up with conventional techniques or at least come close to them in terms of quality. At the same time, it is suggested that in machine learning, the completeness and information content of the base data can lead to a reduction in system efficiency.

2. Advertising and data-financed business models: The number of business models for online activities - which are mainly affected by the new GDPR - is limited (see for example Croll - Yoskovitz (2013)). Part of it uses advertising revenue or sells data to cover running costs and finance innovation activities. It can be assumed that due that the now required explicit consent for the transfer of data, that this will be possible less often than is currently the case. The same applies when it comes to the tracking of users in order to obtain data for the personalised placement of advertisements on websites.

Investigations have shown that current European privacy regulation reduces the efficiency of advertising. This trend will be further reinforced by the upcoming GDPR. It can therefore be assumed that advertising-financed business models will be less suitable for generating sufficient resources. However, this does not mean that advertising on the Internet is a thing of the past. Furthermore, it is possible to advertise efficiently on specialised sites, because the interests of the visitors are known there, or because users have explicitly given permission to use their personal data. Of course, advertising providers will also be looking for ways to adapt to the new situation and further develop their business models. Even now - in spite of the stricter European data protection regulations - the large, mainly advertising-financed American platform providers are also flourishing in Europe. This will not change in the foreseeable future.

It remains to be seen whether, in the future, it will be easier for users to track the flow of their personal data and whether they will pay with their data for free products. In any case, it can be assumed that the GDPR business models with advertising and data-financed innovation activities are less attractive.

Theoretical modelling of this situation clearly shows that providers tend to offer payment models under these conditions and do not rely on products financed by advertising.

3. Data protection as part of the marketing strategy: Even now, users do not want their data to be traded, or used for the purpose of targeted advertising. The vast majority is in favour of strict data protection and control over the dissemination of their personal data. Only a small minority is prepared to pass on personal data.

The often observed privacy paradox - that even users who want to protect their personal data, yet readily exchange it for a free product at the next best opportunity - may be seen as a contradiction to the desire for data protection. However, it should be noted that it is generally very difficult to realistically assess the consequences of agreeing to terms of use of online services. Therefore, users try to make situation-specific decisions that are hardly rational in view of the scarcely available information on data transfer. In many cases, you can only choose whether you want to use a service and therefore have to accept the loss of control over personal data. Particularly in the case of services with strong network effects such as social media, the pressure to comply is particularly strong. At the same time, of course, these providers also try to impart that they handle data with care.

In view of this starting position, the new GDPR will bring improvements that are in the consumer's interest, because it must be made clearer how the data are to be handled, and explicit and informed consent must be obtained. In an environment where both consumers and legislators are pushing for more data protection, it is obvious that making data protection an integral part of the product range will allow businesses to actively differentiate themselves from "data-hungry" providers. Some suppliers have already begun to follow this path.

With the entry into effect of the GDPR in May 2018, every processor of personal data will be forced to make adjustments which emphasise that data protection regulations should be taken seriously, and implement them accordingly. This enables one to make this part of one’s marketing strategy.

Generally speaking, this seems to be an expandable strategy for the development of digital products and services in Europe, which should be pursued at all policy levels, and become a constituent element of horizontal industrial policy. This, of course, does not only include the corporate sector, but should also be the guideline for the public sector - including intelligence services. This could make Europe synonymous with strong protection of personal data. A position that neither Asian suppliers nor the USA are aiming for.

4. Legal requirements allow the development of data protection technologies: In principle, GDPR creates a "market" for data protection technologies (Privacy Enhancing Technologies - PET). These can be requested by those affected (VPN, TOR etc.), as well as by companies that want to guarantee data protection for their customers. The legal framework conditions are essential for the development of data protection technologies in particular, because they create demand for certain products and services.

However, the legal provisions must be sufficiently clear to stimulate both demand and supply. If goals leave too much room for interpretation, only uncertainty increases, and not the number of solutions to achieve these goals. In concrete terms, there is a great deal of room for interpretation when it comes to the requirements for "deleting data" or "transparency" or "data minimisation". In this case, it is most probably only through the courts that the terms will be clarified.

It does not mean that one should commit oneself to certain technologies when it comes to addressing uncertainties in the interpretation of legal requirements. The targets should be technology-neutral and thus only specify clearly what is to be achieved, and not through which technology the targets are to be achieved.

- The GDPR is, of course, also a potential driver for the development of new technologies that make it easier to comply with the regulations, or for which GDPR has only created a market. From the current perspective, however, there are some limitations:
- The postulated principles are partly contradictory. For example, the demand for transparency conflicts with the right to forget. Transparency is not only the subject of numerous industry-specific regulations (Basel 2, SOX, HIPAA), but also arises from the GDPR itself. Furthermore, a considerable amount of time and effort is also involved, and there are various technical options for implementing compliance with the transparency requirements stipulated in the GDPR, and the associated storage of corresponding records for the use of personal data.
- Many provisions are not really operationalised. The selection of concrete security parameters for anonymization, especially the security factor "k" in the context of k-anonymity or related procedures. The same applies to the use of Differential Privacy, where the choice of the factor

Epsilon has yet to be clarified. In the case of data perturbation, i. e. the intersection of real data sets with synthetic data, it must be clarified from which relationship between real data and perturbation data the privacy of the persons involved is maintained.

Clarification of these issues will largely be carried out by courts, although not only the outcome but also the time of clarification is still completely open.

All in all, these uncertainties mean that GDPR will not lead to market formation because it is not possible to predict the direction in which the market will develop. On the other hand, it is clear that there is still a relatively substantial need for research on many issues, and that steps should be taken at the same time to reduce uncertainties through recommendations for implementation.

5. L**ess leeway for process innovations**: Lower efficiency in internal business processes - a significant impact of more data protection according to Goldfarb and Tucker (2012) - by preventing process innovations is another chain of action. Personal data will often not only be used for product innovations, but will also make it possible to redesign and optimise internal processes. This includes advertising and marketing measures as well as accounting. Since many of these services are also provided by third parties, there may well be restrictions on data transfer (Goldfarb - Tucker, (2012)). This applies in particular if the data protection regulations of third party providers contain provisions that provide for the transfer of data.

4. Economic policy recommendations

4.1 Initial situation

With the introduction of the GDPR in May 2018, it becomes clear that European organisations affected by them no longer have the luxury to ignore data protection and the resulting restrictions, irrespective of their effects on innovation. However, this also applies to providers outside Europe, who nonetheless serve European customers.

The interrelationships between data protection and innovation that have been analysed so far show that this is not a direct and linear relationship, but rather that effects occur at different levels that allow different scenarios depending on the feedback from politicians, companies and users.

In concrete terms - and in a somewhat simplistic manner - no direct negative correlation between innovation, big data and data protection is forecasted, because companies may adapt to the new realities by taking data protection and privacy seriously and building up a relationship of trust with their data providers. Then users will be prepared to grant access to personal data. If innovators do not succeed in building a trustworthy environment, consent to use of personal data is likely to be given much less frequently.

Based on the European situation, the following conclusions can be drawn from this:

Stricter data protection regulations vis-à-vis the United States support a stronger position in the area of payment services, as opposed to data monetisation - i. e. the sale of data - because this requires explicit consent, which can also be revoked, and the message of better data protection is given additional credibility by the company's location.

However, the opportunities for payment services with improved data protection in the private sector can be regarded as limited, due to the privacy paradox and the small market size. The lack of transparency about the further use of data exacerbates the privacy paradox. Only if data providers have certainty about how their data will be used, and how they can prevent the data from being passed on, and have alternatives - payment services -, can they evaluate the advantages and disadvantages.

Experience to date has shown that the privacy paradox[8], coupled with attempts by established platforms to position themselves as companies that value data protection, has not harmed the data business. Companies such as Google or Facebook are responding to this with the creation of walled gardens - an ever more comprehensive ecosystem that offers diversified services and thus increasingly extensive access to customer data (Kelley et al, (2010)).

At European level, alternative payment services would only have a chance of success if a digital single market and massive support were available for the build-up of network effects, as well as increased monitoring of compliance with the provisions of the GDPR (see below in the section on measures).

European regulators are increasingly assuming that the exploitation of data by the dominant American platforms may lead to abuse of market power. This is a circumstance which the European competition authorities increasingly want to address (see Scott - Hirst (2017)), and which would tend to open up new opportunities for alternative suppliers. However, new, smaller providers are only marginally more trustworthy for many customers than the big established companies, so far as data protection is concerned. A number of measures are needed to strengthen the credibility of new providers (see below in the measure section).

There are better conditions for European suppliers in the business and public sector. Due to the increase of the penal framework within the framework of the data protection basic regulation and the potential loss of reputation in case of violation of the GDPR, the willingness to pay for data protection and data monetisation capabilities outside the application context are often not given. Similarly, consumer sensitivity to data protection in eGovernment is likely to be higher than in other areas.

Persons responsible for the processing of personal data in Europe must obtain relatively detailed consent for all planned processing steps, or have to make the data anonymous and then carry out their Big Data evaluation strategies. Since there is uncertainty about the method to be used for anonymization, the implementation of which is challenging and can lead to a large loss of data information content , there is essentially only one option left for European companies: as a location and as a company, to rely heavily on the restrictive data protection regulations in Europe in order to win the confidence of consumers in Europe and elsewhere, and thus to become competitive.

4.2 Measures

If we follow the analysis to this point, we need a series of measures at enterprise level, in technology, innovation and industrial policy. Of course, this also includes the public sector.

[8] As regards the willingness to pay for data protection of private customers, there are a number of investigations that deal with the so-called. Privacy policy. This phenomenon refers to the empirical observation that customers who state in a personal conversation that they attach great importance to data protection are then prepared, in a concrete situation, to make their data available to a provider in return for a relatively small consideration, in which, for example, online data protection regulations are agreed without prior reading (Kokolakis, (2015); Kübler, (2011)).

4.2.1 Measures at company level

In general, more awareness is needed among companies. Only some of the changes required by the GDPR and the strategy resulting from this European legislative trickled down to the company level. Accordingly, two messages should be communicated:

1. concrete guidelines as to what measures companies should take.

2. which strategies for dealing with data are promising.

All considered, it is to be expected that European companies with Big Data ambitions will gain the necessary insights not so much through novel technologies, than through changes in how they present themselves on the market, which will lead to building a relationship of trust with their customers, thereby compensating for the more restrictive European approach.

If one abuses the built-up trust, then the data subjects will be much more cautious about giving explicit consent to the processing of their data. Of course, data protection authorities must be given sufficient resources in order to be able to tackle violations and respond promptly to reports of such.

This creates opportunities for globally active European suppliers who offer specialised solutions with a proven high level of protection for these customer groups. One example of such a strategy is Fabasoft, which is the first company to receive the highest 5-star certification for its cloud services according to the international "EuroCloud Star Audit" (ECSA V3.0) and is certified according to ISO 27018 for the protection of personal data. This international standard formulates data protection requirements for cloud providers. They must provide extensive notification, information, transparency and verification obligations in order to create confidence among customers and authorities regarding the processing of personal data in the cloud[9].

The PIA (Privacy Impact Assessment) Framework (Oetzel, Spiekermann, (2012)) is a framework that helps companies proactively integrate data protection into the design and adaptation of RFID products and services. This framework is based on risk assessment methods and provides a structured method including process and document templates. The basic idea behind these approaches is to anchor privacy in the architecture of the offer, and not only in policies.

Data protection must be implemented in all company activities. A close cooperation between management, developers and organisers is therefore necessary: "Privacy by Design is designed as an engineering and strategic management approach that commits to selectively and sustainably minimise information systems, privacy risks through technical and governance controls" (Spiekermann, (2012)). The companies themselves are better protected against hacker attacks and data leaks.

Economic policymakers should pay attention to shaping the process of conversion to the GDPR and provide action-relevant information on how to deal with its challenges. The objective is to support companies in the timely and efficient implementation of the GDPR, so that a relationship of trust can be established and the prerequisites for data-based innovation strategies can be laid down.

This includes instructions on how to design and implement the explicit consent, as well as information about necessary transparency. Here, it is above all the stakeholders who need to step up the pace, and

[9] https://www.fabasoft.com/de/group/transparenz/sicherheit-datenschutz

- only if they do not react appropriately - the public sector. In addition, companies and start-ups can be inspired to cooperate with selected customer segments and thus support development work. These test users are often prepared to deal with the pitfalls of products that are not yet fully developed, and to provide their data for further development.

In concrete terms, attention should be paid to the forthcoming changes and challenges, and strategies for the implementation of the GDPR should be communicated. The following steps are relevant for companies, so that - as already mentioned - data protection is implemented in all areas:

1. Data inventory
    a. Which data applications are implemented within the company?
    b. What data exists in the company?
2. Sensitivity analysis
    a. Is personal data stored/collected/processed?
    b. Is there any other sensitive data in the company?
3. Setting up an orderly process
    a. When planning new data applications.
    b. When new data sources are opened up or when existing data channels change
4. Reduction
    a. Is all that data really needed by the applications? Which can be omitted or not collected at all?
    b. Conversion of processes if sensitive data is collected/processed that is not really required
5. Application of protective mechanisms (iterative process)
    a. Analysis of whether data is anonymised as a complete data record or whether only evaluation results are required.
    b. Choice of anonymization paradigms (e. g. k-anonymity) with suitable parameters.
    c. Quality control - Is the quality of data evaluation with anonymised data sufficient?
    d. Te chnical optimisation of the anonymization processes - e. g. outlier removal
    e. Establishment of a secure analysis environment with strict logging, separation of data spaces, strong access rights concept, deletion of data after processing
6. Deletion and transparency
    a. Creation of a processing model - Which data flows (also aggregated) into which results?
    b. Implementation of suitable mechanisms to trace which data record was used where - this enables direct access for information and deletion.
    c. Where technically possible: Creation of processes and technical tools for the physical deletion of data, e. g. by overwriting. Where this is not possible, documentation where and why.

From the current perspective, the switch to GDPR is likely to be acknowledged only in part by companies, and even then, the extent of adaptations is not clear. This lack of knowledge should change policy as quickly as possible through targeted campaigns.

Of course, all organizations that process personal data will try to signal that they are complying with all laws. It's hard for outsiders to check up on this. In this case, the introduction of a fast certification process could help to establish a quality standard.

4.2.2 Industrial policy

Given the current market conditions, some companies are able to develop a dominant position and thus build up market power. This in turn helps these companies to collect even more data and thus consolidate or expand their market position. Google, Facebook, Amazon are the most important examples of this development. Strict data protection rules could result in these companies no longer being able to fully exploit their data, which should allow smaller competitors to create competitive bids. However, there is no empirical evidence supporting this hypothesis.

It is unlikely that the first mover effects generated by these companies will actually be mitigated by the new GDPR. It is more likely that companies will be able to access the data with the explicit consent of the users and that strict data protection regulations will therefore hit newcomers to the market particularly hard. Campbell et al. (2011) show that the latter in particular could be the case because large, established companies tend to retain user confidence when it comes to compliance with data protection standards. The strict regulation of credit cards in New Zealand is an empirical example of this: shoppers felt that only large incumbent operators were able to comply with the data protection guidelines, avoiding new and small providers.

This line of argumentation applies above all to products aimed at consumers. Indeed, it is difficult to see that the incumbent operators are losing market power under stricter data protection rules and are therefore able to enter new operators. Nevertheless, users in this segment should also benefit from improved data, even if the market power of the quasi-monopolists is not really curbed.

If GDPR is indeed an effective instrument to prevent unwanted data transfer (also for advertising purposes) - and thus also mitigates the privacy paradox - then there should be significantly more payment offers and less "data for product use exchange transactions".

4.2.3 Technology and Innovation Policy

The hypothesis that data protection laws stimulate innovations in providers of data protection technologies (such as Privacy Enhancing Technologies (PET), e.g. encryption) remains valid even if the framework conditions are not favourable in Europe at present. The uncertainties and conflicts of interest within the scope of the GDPR are large and therefore it is not clear in which direction the technologies can or should develop. This uncertainty is caused by the wording and requirements of the GDPR and can also be corrected or limited by further clarification by the relevant authorities.

An essential task for the public sector is to eliminate the uncertainties that GDPR itself brings with it. Although much has to be put into law, there are other ways to deal with the uncertainties. This includes further explanations or the option of discussing different approaches with the authorities in advance. However, this requires the willingness of authorities to provide the service, and to possess appropriate resources to provide this service. If uncertainty is reduced successfully, data protection technology developers will have more incentives to invest in innovative products and services, because a market was created by these regulations.

This approach must be consistently integrated into a horizontal strategy that includes the following areas:

- **Education:** a competitive advantage through better data protection can only be achieved if users are aware of it. For this purpose, awareness must be established in the general public and applied in education. Data protection aspects should play an important role in the digitisation strategy "School 4.0". Appropriate training and further education should also be provided.
- **Research:** This study presents a series of research questions that will enable the efficient implementation of GDPR. These should be integrated into basic research and applied research (e. g. preservation of information during anonymization, ledger architectures to create transparency, resolving the contradiction of the right to be forgotten and traceability).
- **Funding:** In addition to the corresponding research funding, e. g. through the FFG, start-ups offering corresponding solutions must also be supported. In particular, a connection to Blockchain activities (https://www.blockchain-austria.gv.at/, Blockchain Village) has to be established here, especially as this technology provides a fundamentally different basis for the processing of personal data.
- **Legal framework conditions**: As shown above, an adequate interpretation of the consent rules is crucial for the efficiency of Big Data applications. A description that is too detailed requires ongoing adaptations of the user's consent, and leads to incomprehension. Clearly understandable and comprehensible wording should therefore be used for the declaration of consent.
- Establishment of a "MyData Local Hub" in Austria: Membership in the MyData project can support several of these measures:
    - Public sector providers can contribute their expertise as trusted entities within the framework of a public-private partnership.
    - The users become aware of which of their data is stored where. This promotes the responsible handling of consent, while at the same time reducing the trust problems of smaller providers.
    - The local software community can participate in the development through the open source framework and remains motivated.
    - The findings can also be used for the further development of eGovernment.

4.2.4 Geopolitical

Data privacy regulations are desirable from a socio-political point of view, interesting from an industrial policy point of view, and should be indispensable as part of a geopolitical strategy. Europe has misjudged digital technologies as ordinary generic technologies and underestimated the wider economic and geopolitical implications. Because of its ties to the United States, the constant pressure to remain competitive, and the attempt to regain already lost ground, Europe has put its money mainly on the rapid diffusion of digital technologies, at the expense of their appropriation. Instead of a strategic approach - such as China or Russia - a laissez faire approach was chosen.

Developments in recent months have shown that this positioning is not sustainable and that an independent and Europe-centred approach is needed. Europe is not represented in many of the hot topics surrounding digitisation (e. g. AI in the military sector). One cannot afford to remain in this vacuum in the long run, since it is not possible to participate in developing the framework without such hands-on expertise. In order to make progress in this area, the European dimension must be significantly strengthened on these issues. This is difficult, but the framework conditions are much better with a functioning Franco-German axis, as compared to the situation as it was only recently. The Austrian positioning beyond the current topics is not visible here. It would be desirable to actively strengthen the European dimension and thus participate in geopolitical repositioning.

It is clear that the current steps at European level - the pursuit of abuse of market power, compliance with data protection rules - are defensive strategies and, as such, will be important but successful only if active elements are incorporated. GDPR can become an asset if it is also supported by other policy measures in the sense of a horizontal strategy, and if users again have control over their data. The best way to do this is through a system where users can centrally manage access to their data and all queries are documented. To this end, uniform standards and APIs for Europe must be developed and put live as quickly as possible.

On the whole, however, Europe also needs to position itself as a model for effective data protection, if economic opportunities are to be realised. This also means that people who want to "enjoy" the European level of data protection will be granted a kind of e-residency as Estonia is willing to do. Austria could follow suit and become the precursor for a comprehensive European solution. Ultimately - as already mentioned - all companies must comply with the GDPR if they have European customers. If you increase the number of data protection for Europeans beyond eResidency, then European data protection law spreads beyond Europe and can thus become the standard. One can also consider forming alliances with countries between the geopolitical power blocs (e. g. South America, Japan) to adopt the European data protection rules[10].

Finally - and in terms of innovation - the GDPR was not designed to prevent innovation, but to improve data protection. This is also an innovation that entails considerable costs for organisations who have to implement it. Therefore, all accompanying measures should now be put in place to ensure that the changeover is as simple as possible and that it is a success. This includes information campaigns for organisations who are obliged to comply with it, as well as for users. Information education is also necessary for consumers so that they can accept and deal with the new possibilities. The frequently observed data paradox can only be avoided if informed users make use of the new possibilities.

GDPR imposes restrictions on Big Data strategies when personal data is involved. This does not apply if company data or usage data are concerned. It is also clear that in innovation and big data analysis it is not possible to anticipate which data will be necessary. The ball is clearly in the court of those organisations that want to process personal data. Only if the supplier of data considers the receiver to be a trustwhorthy organisation, generous declarations of consent that underpin Big Data and innovation strategies will be accepted.

However, Europe is also called upon to comply with the strict data protection regulations in agreements with third countries. This applies, of course, to the GDPR itself, but also to the repealed Safe Harbour Agreement with the USA and other data exchange agreements (e. g. passenger data) where the previous provisions went far too far. Europe must be consistent in all respects, abide by its own laws at all levels and thus preserve its autonomy.

[10] This would be much more sensible than allowing the exchange of data via bi- and multinational trade agreements, as is currently often envisaged.

5. References

Accorsi, R., On the relationship of privacy and secure remote logging in dynamic systems. In IFIP International Information Security Conference, 2006

Acquisti, A., Adjerid, I., Brandimarte, L., "Gone in 15 seconds: The limits of privacy transparency and control." IEEE Security & Privacy 11.4 (2013): 72-74.

Acquisti, A., Taylor, C., Wagman, L., the Economics of Privacy, Journal of Economic Literature, Vol. 52, No. 2, 2016.

Acquisti, A., H. R. Varian (2005). Conditioning prices on purchase history. Marketing Science 24 (3), 367{381.

Adams, T., Surge Pricing Comes to the Supermarket, Guardian, 4. Junie 2017, https://www.theguardian.com/technology/2017/jun/04/surge-pricing-comes-to-the-supermarket-dynamic-personal-data.

Albrecht, J. P., & Jotzo, F. (2017). Das neue Datenschutzrecht der EU. Grundlagen, Gesetzgebungsverfahren, Synopse, Baden-Baden.

Arieli, D., Predictably Irrational, The Hidden Forces That Shape Our Decisions, Harper Collins Publisher, 2010.

Article 29 Data Protection Working Party (2004), Opinion 10/2004 on More Harmonised Information Provisions: Adopted on 25th November 2004. 11987/04/EN. Available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp100_en.pdf.

Bartolini, C., Muthuri, R., Cristiana, S., "Using ontologies to model data protection requirements in workflows", 2015.

Bellare M. and B. Yee. Forward integrity for secure audit logs. Technical report, Technical report, Computer Science and Engineering Department, University of California at San Diego, 1997.

Bernd, M., Kieseberg, P., Weippl, E. R., Holzinger, A., "The Right to Be Forgotten: Towards Machine Learning on Perturbed Knowledge Bases," in International Conference on Availability, Reliability, and Security, 2016.

Blank, S., Dorf, B., The Startup Owner´s Manual, K&S Ranch Press, 2012.

Blättel-Mink B., Menez R., Kompendium der Innovationsforschung, 2015.

Bonatti, P. A., & Olmedilla, D., Rule-based policy representation and reasoning for the semantic web. Proceedings of the Third International Summer School Conference on Reasoning Web, RW'07, pp. 240-268. Springer-Verlag, Berlin, Heidelberg, 2007.

Bonatti P., De Capitani di Vimercati, S., Samarati, P., An algebra for composing access control policies. ACM Transactions on Information and System Security (TISSEC) , 5(1), 2002.

Bonatti P., Kirrane S., Polleres A., and Wenning R. Transparent personal data processing: The road ahead. In TELERISE: 3rd International Workshop on TEchnical and LEgal aspects of data pRIvacy and SEcurity @ SAFECOMP2017, Trento, Italy, September 2017. to appear

Bradshaw, J. M., Dutfield, S., Benoit, P., & Woolley, J. D. (1997). Software agents. MIT Press, Cambridge, MA, USA, Ch. KAoS: Toward an Industrial-strength Open Agent Architecture, pp. 375–418.

Cadwalladr, C., The great British Brexit robbery: how our democracy was hijacked, Guardian, 7. 5. 2017, https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Campbell, J. D., Goldfarb, A., Tucker, C., Privacy Regulation and Market Structure. mimeo, University of Toronto, 2011.

Casadesus-Masanell, Ramon, and Andres Hervas-Drane, "Competing with privacy." Management Science 61.1 (2015): 229-246.

Chesbrough, H., Open Innovation, The New Imperative for Creating and Profiting from Technology. 2003

Chesbrough H., Brunswicker S., Managing Open Innovation in Large Firms, Survey Report, Executive Survey on Open Innovation, Fraunhofer 2013.

Croll, A., Yoskovitz, B., Lean Analytics, Use Data to Build a Better Startup Faster, O´Reilly, Sebastopol, 2013.

Determann, L., Adequacy of data protection in the USA: myths and facts. International Data Privacy Law 2016; 6 (3): 244-250, 2016.

Dwork, C., "Differential privacy: A survey of results." In International Conference on Theory and Applications of Models of Computation, pp. 1-19. Springer Berlin Heidelberg, 2008.

Evans, P. C., Gawner , A, The Rise of the Platform Enterprise, A Global Survey, The Center for Global Enterprise, 2015, http://thecge.net/wp-content/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf

Fassauer, R., "Personalisierung im E-Commerce–zur Wirkung von E-Mail-Personalisierung auf ausgewählte ökonomische Kennzahlen des Konsumentenverhaltens." (2014).

Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E. & Wagner, D., Android Permissions: User Attention, Comprehension, and Behavior. Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS 2012. ACM, 2012.

Fernández J., Kiesling, E., Kirrane, S., Neuschmid, J., Mizerski, M., Polleres, A., Sabou, M., Thurner, T., Wetz, P., Propelling the Potential of Enterprise Linked Data in Austria: Roadmap and Report. edition mono/monochrom, Zentagasse 31/8, A-1050 Vienna, Austria, December 2016.

Fernández J., S. Kirrane, A. Polleres, and S. Steyskal, Self-enforcing access control for encrypted RDF. In Proceedings of the 14th European Semantic Web Conference (ESWC2017) , Portorož, Slovenia, May 2017. URL http: //polleres.net/publications/fern-etal-ESWC2017.pdf .

Fernández Garcia, J. D., Umbrich, J., Knuth, M. Polleres, A., Evaluating query and storage strategies for RDF archives. In 12th International Conference on Semantic Systems (SEMANTICS) , ACM International Conference Proceedings Series, 2016.

Fruehwirt, P., Huber, M., Schmiedecker, M., Weippl, E. R., "InnoDB Database Forensics," in Proceedings of the 24th International Conference on Advanced Information Networking and Applications, 2010.

Fruehwirt, P., Kieseberg, P., Schrittwieser, S., Huber, M., Weippl, E.R., "InnoDB Database Forensics: Reconstructing Data Manipulation Queries from Redo Logs," in The Fifth International Workshop on Digital Forensics (WSDF), 2012.

Fruehwirt, P., Kieseberg, P., Schrittwieser, S., Huber, M., Weippl, E. R., "InnoDB Database Forensics: Enhanced Reconstruction of Data Manipulation Queries from Redo Logs," Information Security Technical Report (ISTR), Special Issue: ARES, 2013.

Fruehwirt, P., Kieseberg, P., Krombholz, K, Weippl, E. R., "Towards a forensic-aware database solution: Using a secured database replication protocol and transaction management for digital investigations," Digital Investigation, vol. 11, pp. 336-348, 2014.

Fruehwirt, P., Kieseberg, P., Weippl, E. R., "USING INTERNAL MySQL/InnoDB B-TREE INDEX NAVIGATION FOR DATA HIDING." In IFIP International Conference on Digital Forensics, pp. 179-194. Springer International Publishing, 2015.

Goldfarb, A. and C. Tucker (2011a). Online display advertising: Targeting and obtrusiveness. Marketing Science 30 (3), 389{404.

Goldfarb, A. and C. Tucker (2011b). Privacy regulation and online advertising. Management Science 57 (1), 57{71.

Gottlieb, D. and K. Smetters (2011). Grade non-disclosure. Available at NBER: http://www.nber.org/papers/w17465.

Gross-Amblard, D. (2003). Query-preserving watermarking of relational databases and XML documents. SIGART Symposium on Principles of Database Systems

Gürses, S., del Alamo, J. M., "Privacy Engineering: Shaping an Emerging Field of Research and Practice." IEEE Security & Privacy 14.2 (2016): 40-46.

Cadwalladr, C., The great British Brexit robbery: how our democracy was hijacked, Guardian, 7. 5. 2017, https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy

Hansen, M., "Data Protection by Design and by Default à la European General Data Protection Regulation." Privacy and Identity Management. Facing up to Next Steps. Springer International Publishing, 2016. 27-38.

Hayek, F., The Use of Knowledge in Society, American Economic Review, 1945, https://assets.aeaweb.org/assets/production/journals/aer/top20/35.4.519-530.pdf

Hedbom, H., Pulls, T., Hjärtquist, P. & Lavén, A. (2009). Adding secure transparency logging to the prime core. Privacy and Identity Management for Life, pp. 299-314. Springer Berlin Heidelberg

Hippel E., The Sources of Innovation. 1988

Holt J. E. Holt. Logcrypt: forward security and public verification for secure audit logs. In Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54 , 2006.. Logcrypt: forward security and public verification for secure audit logs. In Proceedings of the 2006 Australasian workshops on Grid computing and e-research-Volume 54 , 2006

Holtz, L. E., Zwingelberg, H., & Hansen, M. (2011) Privacy Policy Icons. Privacy and Identity Management for Life, pp. 279-285

Hope-Bailie, A., Thomas, S., Interledger: Creating a standard for payments. In Proceedings of the 25th International Conference Companion on World Wide Web , 2016.

Huizingh E., Open innovation: State of the art and future perspectives. In: Technovation 31, 2011, 2-9.

Jahnel, D. (2010). Handbuch Datenschutzrecht. Auflage, Salzburg.

Kagal, L., & Finin, T. (2003). A policy language for a pervasive computing environment. Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks, pp. 63-74. IEEE Comput. Soc.

Kelley, P. G., Cesca, L., Bresee, J., Cranor, L. F., Standardizing privacy notices: An online study of the nutrition label approach. Mimeo, Carnegie Mellon University CyLab CMU-CyLab-09-014, 2011.

Kieseberg, P., Schrittwieser, S., Schmiedecker, M. Huber, M., Weippl, E. R., "Trees Cannot Lie: Using Data Structures for Forensics Purposes," in European Intelligence and Security Informatics Conference (EISIC 2011), 2011.

Kieseberg, P., Schrittwieser, S., Mulazzani, M., Echizen, I., Weippl, E. R., "An algorithm for collusion-resistant anonymization and fingerprinting of sensitive microdata." Electronic Markets 24, no. 2 (2014): 113-124

Kieseberg, P., Malle, B., Fruehwirt, P., Weippl, E. R., Holzinger, A., "A tamper-proof audit and control system for the doctor in the loop," Brain Informatics, pp. 1-11, 2016

Kokolakis, S., "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon." Computers & Security, 2015.

Kübler, D., Privates öffentlich: Datenschutz ist dem Schnäppchenjäger nichts wert. In: WZB-Mitteilungen, 132, pp. 10-11, 2011 URN: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-3087

Kirrane, S., Mileo, A., Decker, S., Access control and the resource description framework: A survey. Semantic Web , 8(2):311–352, 2017.

Langelaar, G., Setyawan, I., & Lagendijk, R. (2000). Watermarking digital image and video data. A state-of-the-art overview. IEEE Signal Processing Magazine, 17(5), 20–46

Leo, H. Digitalisation and innovation – how new technologies can help to overcome economic barriers, paper presented at OSCE conference “Towards the Vision of a Common Economic Space from Vancouver to Vladivostok: Connectivity, Trade and Economic Cooperation”, 15-16 May 2017, Linz, Austria.

Leo, H., Seethaler, U., Schritte zur Operationalisierung der Open Innovation Strategie für Österreich, Wien, 2017.

Leo, H., Palme, G., Volk, E., Die Innovationstätigkeit der österreichischen Industrie, Technologie- und Innovationstest 1990, Wifo, Wien. 1992.

Li, W., Yuan, Y., Li, X., Xue, X., & Lu, P. (2005). Overview of digital audio watermarking. Tongxin Xuebao (Journal on Communications), 26(2)

Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and l-diversity." In 2007 IEEE 23rd International Conference on Data Engineering, pp. 106-115. IEEE, 2007.

Liu, Siyuan, Shuhong Wang, Robert H. Deng, and Weizhong Shao. "A block oriented fingerprinting scheme in relational database." In International Conference on Information Security and Cryptology, (2004): 455-466

Madden, M. Rainie, L., (2015). Americans' attitudes about privacy, security and surveillance.

Martin, K D., Murphy, P. E., "The role of data privacy in marketing." Journal of the Academy of Marketing Science (2016): 1-21.

Ma D., Tsudik, G., A new approach to secure logging. ACM Transactions on Storage (TOS) , 5(1), 2009.

Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., "l-diversity: Privacy beyond k-anonymity." ACM Transactions on Knowledge Discovery from Data (TKDD) 1, no. 1 (2007): 3

Maurya, A., Running Lean: Iterate from Plan A to a Plan That Works, O'Reilly, 2012.

Mazzucato, M., The Entrepreneurial State: debunking public vs. private sector myths, Anthem Press, London, 2013.

McDonald, A. M., Cranor, L. F., "The cost of reading privacy policies." ISJLP 4 (2008): 543.

McRoberts, M., Doncel, R.V., ODRL Version 2.1 Ontology, 2015 Available at: http://www.w3.org/ns/odrl/2/ODRL21.

Newman, D., The Process of Design Squiggle, Central Office of Design.

Noble, Ch. H., Durmusoglu, S. S., Griffin, A., Open Innovation, New Product Development essentials from the PDMA, Wiley, Hoboken, 2014.

Oetzel, M. C., Spiekermann, C., "Privacy-by-Design through Systematic Privacy Impact Assessment-a Design Science Approach." ECIS. 2012.

Paal, B. P., Pauly, D. A., & Ernst, S. (2017). Datenschutz-Grundverordnung.

Peeters R., T. Pulls, and K. Wouters. Enhancing transparency with distributed privacy-preserving logging. In ISSE 2013 Securing Electronic Business Processes. Springer, 2013.

Piller F., Lüttgens D., Pollok P., Open Innovation. Methoden und Erfolgsbeurteilung, in: Wirtschaftswissenschaftliches Studium Zeitschrift für Ausbildung und Hochschulkontakt, 11-2013.

Piller F., Not Invented here. In: Interview von Ramge Thomas, Brand Eins 01-2017, S. 72.

Pulls T., R. Peeters, and K. Wouters. Distributed privacy-preserving transparency logging. In Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society , 2013.

Rainie, L., S. Kiesler, R. Kang, M. Madden, M. Duggan, S. Brown, and L. Dabbish (2013). Anonymity, privacy, and security online. Pew Research Center.

Ries, E., The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses, 2011.

Rinne M., E. Blomqvist, R. Keskisärkkä, and E. Nuutila. Event processing in rdf. In Proceedings of the 4th International Conference on Ontology and Semantic Web Patterns-Volume 1188 , 2013.

Rogaway, P., "A synopsis of format-preserving encryption." In UNPUBLISHED MANUSCRIPT. 2010.

Sackmann S., J. Strüker, and R. Accorsi. Personalization in privacy-aware highly dynamic systems. Communications of the ACM , 49(9), 2006.

Samuel, J., Zhang, B., RequestPolicy: Increasing web browsing privacy through control of cross-site requests. Privacy enhancing technologies. Springer Berlin Heidelberg, 2009.

Schneier B., Kelsey, J., Cryptographic support for secure logs on untrusted machines. In USENIX Security , 1998.

Schneier B., Kelsey, J., "Secure audit logs to support computer forensics." ACM Transactions on Information and System Security (TISSEC) 2, no. 2 (1999): 159-176.

Schumpeter, J. A., Capitalism, Socialism, and Democracy, Harper, New York, 1942.

Scott, M., Europe’s tech ambition: To be the world’s digital policeman, The Continent’s policymakers want to determine how companies and their users behave online, Politico, 8/20/17, http://www.politico.eu/article/europe-tech-ambition-to-be-world-digital-policeman/

Scott, M., Hirst, N., Europe´s next competition clash: Online data, Politico.eu, 8/25/17, http://www.politico.eu/article/europe-competition-google-amazon-facebook-data-privacy-antitrust-vestager/

Schwartz, P. M., Preemption and privacy. Yale Lj, 118, 902, 2008.

Seneviratne, O., Kagal, L., Enabling privacy through transparency. In Privacy, Security and Trust (PST), 2014 Twelfth Annual International Conference on Privacy, Security and Trust, 2014.

Seth, G., Lynch, N., “Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services.” ACM SIGACT News, v. 33 issue 2, 2002, p. 51–59.

Shellshear, E., Innovation Tools, The most successful tools to innovate effectively and cheaply, 2016.

Singla, S., Kumar, R., Kumar, D., Natural Computing for Automatic Test Data Generation Approach Using Spanning Tree Concepts. Procedia Computer Science, 85, 929-939, 2016

Sion, R., Atallah, M., Prabhakar, S., Watermarking relational databases, 2002.

Skopik, F., Settanni, G., Fiedler, R., Friedberg, I., Semi-synthetic data set generation for security software evaluation. In 2014 Twelfth Annual International Conference on Privacy, Security and Trust, Toronto, ON, Canada, July 23-24, 2014, pages 156{163, 2014

Spiekermann, S., "The challenges of privacy by design." Communications of the ACM 55.7 (2012): 38-40.

Som O., Jäger A., Maloca S., Open Innovation - ein universelles Erfolgskonzept? in: Modernisierung der Produktion. Mitteilung aus der ISI-Erhebung. Frauenhofer, Ausgabe 66, 08-2014.

Steyskal, S., Polleres, A., Towards formal semantics for ODRL policies. In 9th International Web Rule Symposium (RuleML2015) , number 9202, pages 360–375, Berlin, Germany, Aug. 2015. Springer. URL http://www.polleres.net/publications/stey-poll-2015RuleML.pdf .

Sweeney, L., "k-anonymity: A model for protecting privacy." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, no. 05 (2002): 557-570.

Sweeney, L., "Comments to the department of health and human services on “standards of privacy of individually identifiable health information”.", 2002.

Taylor, C. R., Consumer privacy and the market for customer information. RAND Journal of Economics 35 (4), 631{650, 2004.

Tikkinen-Piri, C., Rohunen, A., Markkula, J., EU General Data Protection Regulation: Changes and implications for personal data collecting companies. Computer Law & Security Review, 2017.

Tucker, C. E., The economics of advertising and privacy. International journal of Industrial organization 30 (3), 326{329, 2012.

Turow, J., J. King, C. Hoofnagle, A. Bleakley, and M. Hennessy (2009). Americans reject tailored advertising and three activities that enable it. Working paper.

Vayena, E., Mastroianni, A., Kahn, J., 2013. Caught in the web: informed consent for online health research. Sci Transl Med, 5(173), p.173fs6.

Licenses Compatibility and Composition in the Web of Data. Proceedings of Third International Workshop on Consuming Linked Data, COLD, 2012.

Wagner P., Piller T., Open Innovation - Methoden und Umsetzungsbedingungen, in: Howald J. (Hg.) et.al., Innovationsmanagement 2.0, 2011.

Weitzner D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., Sussman, G. J., Information accountability. Communications of the ACM , 51(6), 2008.

Wolff, H. A., & Brink, S., Beck'scher Online-Kommentar Datenschutzrecht. Aufl. München: CH Beck, 2013.

Wouters K., K. Simoens, D. Lathouwers, and B. Preneel. Secure and privacy-friendly logging for egovernment services. In Availability, Reliability and Security, 2008. ARES 08. Third International Conference on Availability, Reliability and Security, 2008.

Xiao, X., Tao, Y., "M-invariance: towards privacy preserving re-publication of dynamic datasets." In Proceedings of the 2007 ACM SIGMOD international conference on Management of data, pp. 689-700. ACM, 2007

Yang, X-C., Liu, X-Y, Wang, B., Yu, G., "K-anonymization approaches for supporting multiple constraints." Ruan Jian Xue Bao(Journal of Software) 17, no. 5 (2006): 1222-1231.

Zarsky, T. Z., The Privacy-Innovation Conundrum, Lewis & Clark Law Review, 2015.

Zuckerberg, M., Building Global Community, 2017, https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/

Zuiderveen Borgeswius, F. J., Security & Privacy, 'Informed Consent. We Can Do Better to Defend Privacy', IEEE (Volume 13, Issue 2, p. 103-107).