# The impact of general data protection regulation on software engineering practices

Luís Leite and Daniel Rodrigues dos Santos
*Gaya Higher Polytechnic Institute, Vila Nova de Gaia, Portugal, and*
Fernando Almeida
*INESC TEC R&D Centre, University of Porto, Porto, Portugal*

Received 28 March 2020
Revised 25 July 2020
20 November 2020
11 March 2021
21 June 2021
Accepted 21 June 2021

## Abstract

**Purpose** – This paper aims to explore the changes imposed by the general data protection regulation (GDPR) on software engineering practices. The fundamental objective is to have a perception of the practices and phases that have experienced the greatest changes. Additionally, it aims to identify a set of good practices that can be adopted by software engineering companies.

**Design/methodology/approach** – This study uses a qualitative methodology through four case studies involving Portuguese software engineering companies. Two of these companies are small and medium enterprises (SMEs) while the other remaining two are micro-companies. The thematic analysis is adopted to identify patterns in the performed interviews.

**Findings** – The findings indicate that significant changes have occurred at all stages of software development. In particular, the initial stages of identifying requirements and modeling processes were the stages that experienced the greatest changes. On the opposite, the technical development phase has not noticeably changed but, nevertheless, it is necessary to look at the importance of training software developers for GDPR rules and practices.

**Research limitations/implications** – Two relevant limitations were identified as follows: only four case studies involving micro-companies and SMEs were considered, and only the traditional software development methodology was considered. The use of agile methodologies was not explored in this study and the findings can only be mainly applied to the waterfall model.

**Originality/value** – This study offers mainly practical contributions by identifying a set of challenges that are posed to software engineering companies by the implementation of GDPR. Through their knowledge, it is expected to help these companies to better prepare themselves and anticipate the challenges they will necessarily face.

**Keywords** Software engineering, Software industry, Data governance, General data protection regulation (GDPR), Privacy compliance, The software industry

**Paper type** Research paper

## Introduction

Organizations are beginning to show a growing concern for information privacy due to a diverse set of factors such as higher customer demands, the evolution of data protection law and the organizations' own proactivity (Baruh *et al.*, 2017; Li, 2011). Cascio and Montealegre (2016) state that the convergence and rapid evolution of technology have transformed the interaction between people and organizations, and have fostered the sharing of corporate and personal information. Social networks, mobile devices and cloud computing have contributed to dissolving the boundaries of organizations. The adoption of new technologies presents not only numerous benefits but also exposes organizations to a set of new threats. At present,

Information & Computer Security
Vol. 30 No. 1, 2022
pp. 79-96

2056-4961
DOI 10.1108/ICS-03-2020-0043

computer attacks are more sophisticated and require less technical knowledge from attackers, namely because the internet provides a set of advanced tools that exploit the weaknesses of new technologies (Bendovschi, 2015).

In the current context, data security is under constant threat. The challenges to cybersecurity are incessant and are permanently more sophisticated, taking the most diverse forms (e.g. denial of service, identity theft, phishing and malware). These security breaches are directed at both data subjects and the organizations that have access to them and result in the disclosure of personal data that can be used for illicit purposes and cause serious harm to data subjects. The average costs associated with data breaches continue to rise from US$3.86m in 2018 to US$3.92m in 2019 (Taylor, 2020).

The issue of privacy is present in all fields of personal life and in citizens' interactions with companies. Although there are several definitions of privacy, this is fundamentally a diffuse concept and can have several interpretations and practical applications. The concept of privacy can relate to the specific aspects of an individual's life, access to which may be allowed to others based on interpersonal relationships and trust; on the opposite side, it can relate to the innermost sphere of the person, i.e. to more restricted and deeper aspects of an individual's life (Mulligan *et al.*, 2016). Privacy is considered a fundamental right of the person, which must be safeguarded and, to this end, instruments for its protection must be ensured.

The general data protection regulation (GDPR) is the new European legal framework that entered into force on May 25, 2018, and aims to regulate the process of collection, processing and management of personal data. The GDPR applies to all European Union (EU) companies and organizations that hold or process personal data. Furthermore, it is also extended to companies in third countries that provide goods or services to citizens in the EU or monitor their behavior there.

The analysis of the impact of GDPR on companies is still a relatively new area. The study conducted by Teixeira *et al.* (2019) summarizes the barriers and enablers for the adoption of GDPR while Poritskiy *et al.* (2019) look specifically at the benefits and challenges raised for the information technology (IT) sector. In both studies, the impact of the GDPR is comprehensively analyzed considering the outcomes of the GDPR for firms in general. These studies look at the GDPR from the perspective of the legal changes that must be accomplished by companies and the outcomes that are generated. However, the practices that had to be changed to engage the GDPR by software engineering companies are not addressed and analyzed. In this sense, this study intends to explore the impact of the GDPR considering the different software engineering practices. The aim is to know the areas that suffered the greatest impact and to collect a set of good practices for software engineering companies to apply in their work processes.

This study is organized as follows: In the first phase, a literature review is carried out in the GDPR field, focusing explicitly on its main principles and practices. After that, the adopted methodology in this study is presented. Next, the results are discussed and compared with the existing literature in this field. Finally, the main conclusions of this study are enumerated and topics for future work are identified.

## Literature review

### *The general data protection regulation*

The right to the protection of personal data derives from one of the fundamental human rights proclaimed in the Universal Declaration of Human Rights (Duan, 2017). Our personal data constitute information from our private life and should, therefore, be protected.

According to Esteve (2017), private life should be viewed as everything that relates to our personal lives, from the most intimate data to the data of our professional and social life.

In the GDPR (Art. 4, n. 1) personal data covers all information relating to an identified or identifiable person. The definition of personal data covers all data relating to any person who is identified, isolated or can be identified or isolated. Therefore, as Crutzen *et al.* (2019) argue, a person can be identified by his name, identification number, taxpayer number, location data, etc. Furthermore, the elements present in the metadata, big data and profile definition are also part of the personal data (Wachter, 2019).

The concept of sensitive data has also been extended in the GDPR to cover biometric data. Special categories of personal data (e.g. ethnicity, political opinions, religion, genetic data) are established in Art. 9 of the GDPR. This position is reinforced in Art. 51 by stating the processing of sensitive data may entail significant risks for fundamental rights and freedoms. Therefore, the processing of sensitive data is prohibited, except where there is the explicit consent of the data owner or processing is necessary for reasons of public interest (Dove, 2018).

Art. 6 of the GPDR presents the conditions for processing personal data. It includes the conditions for the lawfulness of processing, respectively, as follows:

- The consent of the data subject;
- The necessity of the processing to fulfill a legal obligation; or
- The existence of legitimate interests pursued by the controller or third parties.

It should also be noted that for children under the age of 16, treatment must be authorized by the holders of the child's parental responsibilities (Macenaite and Kosta, 2017). This age can be lowered up to 13 years by the decision of the member states.

The purposes for which data are processed must be unambiguous. The purposes of data processing must be explicit and legitimate and must be determined at the time of collection of personal data, as laid down in Art. 39 of the GDPR. Accordingly, as Crutzen *et al.* (2019) also state, the purpose of the processing must be known to the data owner before the processing starts.

Another fundamental principle defined in the GDPR is the minimization of data. According to Art. 39, personal data must be adequate, relevant and limited to what is necessary for the purposes for which they are processed. Therefore, the minimum amount of data must be collected for the purpose and excessive and unnecessary data should not be collected. From this principle, it can be concluded that instead of working with personal data, these data should be anonymized. Anirudh and Massillamani (2019) state that encryption and anonymization ensure the protection of data privacy, regardless of where the data are stored and throughout their life cycle.

The GDPR (Art. 35, n. 1) also establishes the principle of accuracy, i.e. the data collected must be correct and up to date. The rectification of data must be carried out with due speed when they are incorrect or outdated. Furthermore, the data must be kept only for the period necessary for the purposes for which it is processed (Art. 17). Consequently, the data owner may request the deletion of his/her personal data without undue delay and the data controller is obliged to delete the personal data without delay.

Finally, the GDPR also establishes the role of the data protection officer (DPO). In accordance with Art. 37 n. 5, the DPO must be appointed based on his/her professional qualities or based on his/her expertise in data protection law and practice. The DPO's role should focus on the following five essential tasks (HIPAA Journal, 2018):

- To inform and advise on the obligations of the DPO;

- To monitor and control the organization's data processing;
- To monitor the conduct of a data protection impact assessment (DPIA);
- To act as a contact point for requests concerning the processing of personal data; and
- To cooperate with data protection authorities.

*The concept of data protection by design and default*

In connection with the implementation of the GDPR, two terms of great importance have emerged regarding data protection and privacy on the internet. Data protection by design stipulates that security measures and organizational techniques should be adopted to ensure the protection of users' personal data privacy (Schartum, 2016). Accordingly, from the earliest stages of the development of a process involving the processing of personal data, measures must be adopted to ensure the protection of personal data. This protection must be ensured from the moment software is created, planned and developed. As Clearwater and Philbrook (2018) state, privacy should be in control of the development of a project, product or service. This model should serve as a source of inspiration for companies to incorporate privacy among their values and reinforce their commitment to ethics and transparency.

Data protection by design is based on seven principles (Cavoukian, 2006) as follows:

(1) Proactive and non-reactive and preventive, not remedial;
(2) Privacy by default;
(3) Privacy embedded in design;
(4) Full functionality;
(5) End-to-end security;
(6) Visibility and transparency; and
(7) Respect for user privacy.

For these principles to be fulfilled in practice, it becomes essential to have consent regarding the handling of the user's personal data, with correct and updated information. It is also essential to ensure that the user can always have access to their data.

Data protection by default is a consequence of data protection by design. Data protection by default assumes that software when released to the market should come with privacy settings in the strictest possible mode by default (Bygrave, 2017). Only data essential to provide the service or deliver the product should be collected. Furthermore, the user must be informed about the information is being collected and for what purpose. However, the user may release access to the collection of more information if he/she deems it necessary. An example of the practical adoption of data protection by default is given by Wiesemborski (2019), in which he states that cookies from a website should only be enabled when the user voluntarily activates this data collection. GDPR requires all companies that use cookies to leave them disabled by default. The user will be able to decide at a later stage which data he/she wishes to share.

## Methodology

The impact of the GDPR is essentially analyzed in the literature from legal and business perspectives. The analysis of its impact on software engineering practices is an unexplored area and therefore, it is relevant to explore its impact on software engineering companies. In

addition, the impact of the GDPR can cover multiple domains and it is also relevant to frame the context of organizations. In this sense, the authors of this study have chosen the qualitative methodology through the adoption of four case studies. The case-study approach turns possible to assess a given situation within its context and offers a flexible method for data analysis (Queirós *et al.*, 2017). According to Yin (2017), the case study (CS) method can be applied using the following three approaches: exploratory; descriptive; and explanatory. In this study, exploratory research was adopted because the aim was to explore the impact of GDPR's fundamental rights on software engineering practices.

A total of four case studies were considered. The profile of the companies is shown in Table 1. The size of the companies took into account the OECD (2020) framework considering the number of employees, respectively, as follows:

- The micro company has a maximum of 10 employees;
- The small or medium enterprise (SME) has less than 250 employees; and
- The large enterprise has 250 or more employees.

All companies develop products and services in the software engineering field and adopt the waterfall software methodology. Interviews were registered using handwritten notes and audio recordings. The interviews were conducted in person and by Skype with team leaders of software engineering teams during the period from January to March 2020. Each company has appointed a team leader responsible for managing software development projects. The team leader is responsible for setting project priorities and ensuring that the team is constantly focused on providing value to the company. The main difference between a team leader and a project manager is that in the first case there is a focus on the team while a project manager is focused on processes. Both team leaders are involved in the backend and frontend software development processes. The interviews were subsequently transcribed into an individual report which was subsequently validated by each company's team leaders.

A key element throughout this process was the identification of software engineering practices. The framework proposed by Sommerville (2015) was adopted for this purpose, organizing these practices in the following six major areas: a collection of requirements; planning of activities; analysis and modeling; technical development; testing; and deployment and delivery. The approach followed is linear and there is a sequencing of activities.

| CS | Established year | Size | Description |
|---|---|---|---|
| CS1 | 2017 | Micro company | Develops custom software for various platforms and channels. It also offers IT consulting services. It works mainly for the international market |
| CS2 | 2015 | SME | It offers technological solutions for the business market in the textile area. All its clients are international, acting mainly in the European market |
| CS3 | 2007 | SME | Develops software in the information systems field with specialization for the health market. It covers exclusively the domestic market with hospitals, health centers and medical clinics |
| CS4 | 2008 | Micro company | Software development company for the education and vocational training market. There is a modular-based solution that is customized according to the characteristics of each institution |

**Table 1.**
Profile of case studies

## Results

A thematic analysis was adopted to interpret the results of these interviews. A protocol composed of six steps was as follows:

(1) Organization and preparation of the data for analysis, in which the interviews were standardized and any spelling and syntax errors were eliminated;

(2) Compilation of all data from the four case studies;

(3) Specification of the coding process in the webQDA software;

(4) Application of the coding process to generate a description of the scenario;

(5) Prediction of how the description and themes will be represented in the qualitative narrative and for this purpose, the themes identified were mapped in each interview; and

(6) Extraction and interpretation of the meaning of the data.

The data analysis and treatment work were performed using the webQDA software, in which open procedures were used to progressively develop themes based on rigorous data triangulation (Freitas *et al.*, 2019). The most cited words were as follows: data protection (43); information security (41); company/personal data (36); documentation (33); protection by default (27); DPO (24); protection by design (19); user-friendly (16); consent (15); risk assessment (12); incident management (12); data breach (11); test data (9); open standards (8); mitigation plan (8); data encryption (8); requirements (6); and audit (5). Stop words related to articles, conjunctions and verbs were eliminated. These themes were organized according to the area of each software engineering practice. Most of the themes found are common to all four case studies. These situations were identified in Table 2 as "All" and, if specific to each CS, are identified as CSx, in which each "x" represents the number of the CS. However, in the case of the "testing" phase, the themes identified are only supported in CS3, based on the specificity of the development of its solutions for the health sector, which requires a review of their testing process.

The results indicate that in all engineering software practices there have been changes in companies. However, there are stages such as the collection of requirements, analysis and modeling and deployment and delivery where the number of changes caused by the emergence of the GDPR becomes more notorious while the processes related to the technical design of solutions are those that have remained less changed. Furthermore, there is a set of practices that are defined in the initial phases of the software development process like data protection by design and by default that also manifest themselves in other phases of the software development cycle such as analysis and modeling. In the technical development, no change has been reported, beyond the need for training of software developers in the field of security and privacy.

## Discussion

### *Collection of requirements*

In the four case studies, it has been possible to identify the concern of companies in applying Art. 25 of the GDPR, in which data protection is requested by design and by default. Article 25 requires the data controller to apply appropriate techniques and measures to ensure the protection of data owners. This approach should be implemented at an early stage when the requirements of a project are captured. Although these principles are already applied under the data quality principle as highlighted by Monreale *et al.* (2014) and Romero and Heredero (2017), the GDPR expressly insists on them when defining the data processing techniques.

| Software engineering practices | Final themes |
|---|---|
| Collection of requirements | All: data protection by design (FT1) |
| | All: data protection by default (FT2) |
| | CS1 and CS3: map company's data (FT3) |
| | CS1: risk assessment and mitigation plan (FT4) |
| | All: DPO involvement (FT5) |
| Planning of activities | CS2 and CS3: procedures for handling personal data (FT6) |
| | All: obtain informed consent (FT7) |
| | CS1: information security measures (FT8) |
| Analysis and modeling | CS2 and CS3: increased relevance of non-functional requirements (FT9) |
| | CS2 and CS3: the importance of documentation (FT10) |
| | All: user-friendly design (FT11) |
| | All: storage limitation (FT12) |
| | All: data encryption (FT13) |
| | CS1 and CS3: adoption of open standards (FT14) |
| | CS4: data validity (FT15) |
| Technical development | All: formation of software developers (FT16) |
| Testing | CS3: generation of test data (FT17) |
| | CS3: deletion of test data (FT18) |
| Deployment and delivery | CS3: PETs (FT19) |
| | All: data erasure (FT20) |
| | CS3: database audits (FT21) |
| | All: incident management (FT22) |
| | All: notify data breach (FT23) |

**Note:** FT = Final theme

**Table 2.**
Thematic analysis

Data protection by design stipulates that companies must assume a proactive attitude, creating quality and performance standards from the early stages of a project. Consequently, it must be incorporated into the specification of technological solutions, physical infrastructures and business practices. To increase the visibility of this process, CS3 uses a security workflow in which the client from the beginning can follow data handling and security practices throughout its life cycle. The challenge of implementing data protection by design should also be incorporated into the DevOps model. In this model, it becomes even more apparent that data protection by design does not only apply to the software design phase but also must be incorporated throughout the development, testing and implementation process (Guerriero *et al.*, 2017). DevOps can help with GDPR compliance. As long as the compliance requirements are well defined, it is possible to create the corresponding quality control testing to verify these requirements. This can be done by incorporating these quality assurance checks into the DevOps chain of products in the continuous integration and continuous delivery pipelines. In this way, the conformity of the products can be easily measured and controlled. On the opposite side, data protection by default means that the strictest privacy settings will be applied so that essential data is only used for the specific purpose for which it is intended. CS3 reports that to meet this requirement the company must understand its internal IT structures and architectures. CS1 reports that this challenge can be more easily addressed in the context of SMEs where the technological complexity will necessarily be lower, although the means available for this are also more limited. Furthermore, the information collected must be kept only for the time needed to provide the product or service. This poses new challenges for companies, particularly when designing a new product or service and, as Li *et al.* (2019) indicate, may

increase the time needed to develop new solutions. The approach taken by CS1 to minimize the development time of these solutions is to adopt the Scrum methodology in which all user stories are required to explicitly address the challenges posed by GDPR.

Two of the best-known techniques are the minimization of personal data processing and the pseudonymization of personal data (Crutzen *et al.*, 2019). In data minimization, it is intended that unnecessary information is not collected for each data processing activity. Associated with this principle also emerges the concept of proportionality (Guinchard, 2018). This only aims to collect the minimum data that an entity needs for a certain purpose and these should be proportional. CS2 advocates that this principle should be applied simultaneously to clients and employees. When implementing GDPR in 2018 the company felt the need to review its data collection practices. Before the implementation of GDPR, there was a habit of transversally collecting large data sets (e.g. date of birth, address, zip code, marital status) which then fed specific applications from the human resources department, marketing, purchases, etc. This approach was intended to increase the reuse of information between the various departments in the company. This approach necessarily had to be reviewed, which required a great effort from the IT team. The approach taken was to have each department explicitly define its needs so that access to this information is restricted to each application. Finally, pseudonymization is a process in which there is a replacement of direct and indirect identifiers by coded identifiers that do not allow their interpretation outside the context in which they are used (Starchon and Pikulik, 2019). As Neumann *et al.* (2019) state, pseudonymization contributes to increase data security and may decrease the data caused by possible leaks. This advantage was also mentioned by CS1 as pseudonymized data and their encryption processes are kept separate. The approach taken in CS1 is to use a digital secret key that is changed over a period of two weeks. Therefore, in the event of a leak, the data affected are only those that are not identifiable, without access to the complementary data of the coding process that is kept separately.

Art. 30 of the GDPR states that an organization shall keep a record of the activities performed on the data. However, it does not specify how this process should be implemented. In the process of collecting requirements, it becomes essential to identify the origin of the data (e.g. electronic, video, audio, paper) and a first challenge as outlined in the CS1 is to decide what information needs to be kept and in what format. Biscoe (2017) suggests that a data mapping taxonomy consisting of the following four dimensions should be applied, respectively:

(1) Data items (e.g. names, email addresses);
(2) Formats (e.g. database, online data entry);
(3) Transfer methods (e.g. telephone, internal/external); and
(4) Locations (e.g. cloud, third parties).

This approach helps organizations to have a complete and detailed view of how data influence the organization and allows for the preventive identification of gaps that may exist in data processing (Brodin, 2019). Despite the relevance of this process, CS3 reports that mapping this information alone is not enough. CS3 deals with sensitive data, and therefore, they needed to extend this model and identify special or sensitive personal data, such as the type of illness, advised treatments, children's health, etc.

The GDPR requires organizations to conduct a DPIA where potential data processing operations could be considered invasive. For this purpose, in all case studies carried out, it has been possible to identify the existence of risk assessment and mitigation plans. Through this approach, it can be shown that the rich of personal data breaches have been eliminated

or largely mitigated. CS1 states that companies operating in multiple business sectors have a greater challenge, as the risks of these processes will also be more diversified and, even if their impact in one business area is reduced, it has an amplifying effect for other business areas through breaches of trust and credibility. Among the risks identified by CS1 in its activity are excessive or unauthorized data collection, unauthorized access to data, accidental data alteration or destruction, storage of outdated data, use of data beyond what is expected or socially acceptable, etc.

Finally, the DPO's involvement in the data collection process was considered crucial by all four case studies. Despite their relevance, practical challenges need to be considered. The first of these is the background to the DPO's training, each of these organizations has followed different approaches. CS1 chose to have a DPO with a computer science engineering background having attended advanced training in the specific area of GDPR; while CS2 and CS3 chose to have a DPO with specific training in the law field. None of these approaches is immune to practical challenges. CS1 reports difficulties in having an in-depth knowledge of all legal consequences of the decision; while CS2 and CS3 note difficulties in coordination between IT teams and the DPO. The solution to be adopted depends strongly on the profile of the DPO which as stated in Hoofnagle *et al.* (2019) can have a very heterogeneous profile. An essential aspect that has emerged is the DPO's role in protecting data among the organization's employees. As also reported by Nasir *et al.* (2019), a major factor in information security is the creation of a safety culture within the organization. However, organizational culture cannot be imposed and can only be promoted by the organization and results from the individual and team practices of its employees (Chen *et al.*, 2015). It is in this aspect of behavior change that DPO also has a key role in being responsible for training employees in good practices that enable greater knowledge about confidentiality, integrity and data security.

*Planning of activities*

Most of the personal data processed by organizations use IT tools that should be appropriate and ensure a correct application of the GDPR. CS2 reports that the following two principles are ensured in the processing of information:

(1) Security of access to applications through the use of authentication and identification methods; and

(2) Tracking access.

Notwithstanding the relevance of these two requirements in ensuring data access security, the GDPR application has a wider scope. Truong *et al.* (2020) state that to guarantee the rights of the data owner and the principles of treatment of the GDPR it is important to consider the traceability of the information produced and processed throughout its life cycle. CS3 complements this vision by arguing that computer security is a fundamental pillar of their activity and goes much further than RGPD. CS3 gives the example of sharing information over the phone or by a post-it with co-workers which is a basic safety rule, but it is not imposed by the GDPR. Only if there is a leak because of this behavior we can consider a non-conformity covered by the GDPR. In CS3 computer security is based on a multiplicity of pillars such as access to installations or sensitive areas of the installations, control and registration of access to the network or business devices, control and registration of access to network resources or isolated devices and adequate training of employees.

The processing of data based on the data owner's consent requires the demonstration that the data owner has given explicitly their consent to the data processing operation. In planning the activities of a project must be foreseen at which stages this consent will be

requested. Clifford *et al.* (2019) state the declaration of consent must be provided in clear and simple language and without unfair terms. Furthermore, consent is only a legal basis if control is provided to the data owner and a fair option to accept or refuse the proposed terms without being prejudiced. CS2 reports that this situation has changed its development paradigm in the use of checkboxes, which were by default active. With the GDPR, the pre-validity options are not considered as valid consent under the GDPR. CS1 also reports changes in the planning paradigm of a project's activities when developing technological solutions for minors. In these situations, it is necessary to request consent from those legally responsible before requesting consent (Macenaite and Kosta, 2017).

Actions were taken to ensure data security should be documented, as the GDPR imposes a proactive responsibility on organizations to develop technical and organizational measures to ensure that data processing is conducted in accordance with the GDPR. In practice, this implies the existence of measures capable of ensuring reliability, integrity and availability of access to data, in addition to the existence of a process to regularly test and evaluate the effectiveness of technical and organizational measures. One of the challenges highlighted in CS1 was the outsourcing of development services particularly outside the EU market. Art. 28 of the GDPR stipulates that the controller may only use a subcontractor offering sufficient guarantees, which must be established in a written contract between the parties involved. Accordingly, and in line with Art. 28, CS1 creates a data record that helps document the process and monitor its implementation, which is critical to show to the data protection association the actions and their progress in the case of a data breach.

*Analysis and modeling*

Another software engineering phase that has changed is the analysis and modeling. It stands out the greater relevance given to non-functional requirements, particularly in terms of security, privacy and integrity. Furthermore, the GDPR stipulates that information flows must be maintained in the different systems, from data collection, through processing, export to other systems or their destruction. The user experience (UX) is another factor that will also have to be foreseen and that the GDPR affects the level of user consent and the processes of management and deletion of their data. According to Almeida and Monteiro (2017), the quality of UX has a great impact on the results of a business over time, and therefore, it is essential to predict its impact with the adoption of the GDPR. In this sense, the challenge for software engineering companies is to present the legal requirements of GDPR in a simple way that helps users to manage their data. This view is complemented by Barrett (2019) by stating that the GDPR not only affects the way user data is collected and managed but also has a direct impact on user interface design. The challenge for UX designers lies in its ability to make the UX smooth and pleasant while integration GDPR obligations, which can seem cumbersome and restrictive. Therefore, companies need to offer interfaces that are able to collect the user's consent while offering to the user a fluid path. Best practices like are addressed in Bluestone (2021) and Groen (2019) should be adopted (e.g. contact forms, registration, cookie windows).

The pseudonymization of personal data is recognized in the Art. 32 of the GDPR as a way of reducing the risks of exposure of data owners and providing additional security for those responsible for the processing. All case studies report the use of encryption methods (e.g. MD5, MD6, SHA-3) in implementing this process. Pseudonymization does not remove identifying information from data but removes the linkage of a data set to the original identity using encryption. Consequently, the original data remains unintelligible and the process cannot be reversed without access to the correct decryption key. CS3 reports that to pseudonymized a data set efficiently, all additional information must be kept separate and

subject to technical measures to ensure that it is not assigned to an identified or identifiable person. Furthermore, CS1 states that encryption and decryption operations should be performed locally and not in cloud systems to ensure privacy in the processing of information. Moreover, the life cycle of personal data also changed with an associated expiry date. Finally, to perform pseudonymization efficiently, CS1 advises that all additional information should be kept separately and subject to technical measures that ensure that it is not attributed to an identified or identifiable person.

The principle of minimization is enshrined in Art. 39 of the GDPR stipulates that only relevant and strictly necessary data for which they are processed may be processed. In this sense, no more personal data may be collected from the data subjects than is necessary and sufficient for that purpose. This has led to changes in the design and modeling process of a software application, especially in those where there is a strong emphasis on the collection of personal data. Therefore, before starting the development process, it must be clear what personal data will be collected and processed by the application. Additionally, it is also relevant to consider the validity of the data. The purpose of this process is to ensure that the data is correct and up to date. CS4 gives an example of changes that have been put forward by GDPR in the process of registering new trainees. Traditionally information was collected on the affiliation of the trainees. However, nowadays only the employer's value added tax is requested, except in situations where this requirement results from obligations arising from contracts entered into by the processing entity, as in the case of training carried out with EU funding.

Finally, the importance is given to open standards also increases. Interoperability is a key factor in GDPR and for its implementation to be possible or facilitated, it is necessary that the various applications share data using open standards. Consequently, the increase in interoperability mechanisms results in less effort to create interoperation interfaces that enable faster and more agile communication (Reynolds and Wyatt, 2011). Furthermore, Li *et al.* (2019) advocate that open standards can help organizations comply with GDPR while ensuring interoperability and high UX standards. In CS1, it is mentioned that the use of open-source software in most of its solutions is an element that facilitates the adoption of open standards while CS3 states that the existence of standards for information transfer in the medical field was also an element that has made the company to always adopt open standards, as its conception.

*Technical development*

The technical development phase was recognized transversally by the four case studies as the phase that suffered the least direct changes. In the implementation phase, it is necessary, above all, that there is a training of software developers for GDRP rules and practices. CS1 states that this point is a key element to the success of a project. To this end, and considering the limitations of the company's financial resources, the company benefited from courses in the area promoted by its science park. Two courses were attended on GDPR Foundation and GDPR Practitioner. Furthermore, the way the company is organized should allow different access according to needs. Therefore, it must be guaranteed that only those who need it have access to the information.

The change of paradigm imposed by the GDPR implies a profound adaptation of internal processes and systems. Titus (2018) states this challenge can only be met with IT professionals duly aware of the implications of the new legal regime. The levels of preparation for the GDPR in EU members are quite heterogeneous (Tikkinen-Piri *et al.*, 2018). Studies conducted by KPMG in 2017 report that only 10% of companies in Portugal consider that they promote adequate awareness and training actions on personal data protection (KPMG, 2017). The

natural evolution of this indicator is expected as the GDPR is more widely disseminated and applied, although challenges remain in this area. CS3 reports that one of these challenges is due to the high turnover rate of personnel in the IT area, which means that training cycles have to be constant. Furthermore, the technical challenges, particularly with the proliferation of social networks and big data, increase the need for training in these fields.

*Testing*

Testing is another phase in the software engineering pipeline that is also undergoing changes. In case the tests include personal data of EU residents it becomes necessary to ensure that they are collected, processed and stored according to the GDPR. Consequently, the real data used in the testing process must be properly masked. One way to ensure this approach is the adoption of synthetic data (Maynard-Atem, 2019). This process can be decomposed according to Marinina (2019) into the following three phases:

(1) Identifying the location where personal data is stored;

(2) Masking personal information to ensure that the actual person associated is anonymous and not identifiable; and

(3) Limiting the exposure of such information to only those persons who need to use it.

CS3 reports this is a complex and time-consuming process. In this sense, one approach to limiting your exposure and possible security breaches is to use automated tools for generating synthetic data (e.g. DATPROF, DTM Data Generator, Solix EDMS).

The GDPR also states that the data used in the test process must be deleted when they are no longer needed. Therefore, a mechanism should be implemented to ensure the elimination of test data after its completion. Apparently, the definition of this mechanism is technically simple, but it raises other challenges considering the performance of the testing phase. CS3 reports the increased time spent on testing of more than 25% because the test data from other applications could no longer be used. In this sense, the company automated the process of generating synthetic data.

*Deployment and delivery*

Technological and legal developments have led to an increase in awareness of the importance of information security and privacy. Van Dijk *et al.* (2018) state privacy-enhancing technologies (PETs) are an approach to implementing the data protection by design paradigm and contribute to protect privacy, eliminate and reduce the need for personal data. However, PETs are only adopted at CS3. The difficulties for their practical implementation are reported by OPCC (2017), particularly the few commercial incentives and the high complexity of these platforms.

Art. 16 and 17 of GDPR establish the right of rectification and right of erasure. The former establishes the right of the data owner to rectify his/her personal data if it is incorrect, outdated or incomplete; the latter enables the data owner to delete his/her data and implement the "the right to be forgotten" paradigm. Furthermore, Art. 21 of GDPR establishes that the data owner has the right to object at any time to the processing of his personal data. In this sense, even if the personal data are not erased, they may not be used in that context. The complexity of implementing these operations arises essentially from the company's ability to identify and record the processing of personal data. The adoption of PETs as highlighted in CS3 helped the company to deal with these situations. However, their adoption requires verification of the following conditions as stated by Politou *et al.* (2018):

- The data must be unnecessary for the purpose for which they were collected;
- The data owner withdraws his consent;
- The data owner opposes automated processing; and
- Where personal data have been processed illicitly.

CS2 stresses that it is important that actions arising from a data erasure request be reviewed as data stored for compliance with legal and tax obligations are excluded from this right.

The process of deleting data has to be answered within 30 days and involves the adoption of a workflow to ensure compliance, namely:

- Checking the legitimacy of the request;
- Identifying the relevant data to be deleted; and
- Deleting the data.

The GDPR further requires the organization to take all reasonable steps to ensure that personal data are also erased if they have been shared with other third parties or made publicly available. CS1 reports that the greatest challenge lies in erasing data from the public domain, as this involves coordination with external parties who do not always assume a high level of speed in processing these requests and it is extremely difficult to trace all of their uses.

Finally, other elements emerge as relevant in the deployment and delivery phase. Database audits will determine whether the defined policies and procedures are being properly implemented or need to be improved (Hoofnagle *et al.*, 2019). There also arises a need to implement incident management to identify whether the organization's systems or data have been compromised (Li *et al.*, 2019). The organization should also be able to restore access to personal data promptly if any physical or technical incident arises. All four case studies have incident management platforms and this is particularly relevant to CS3 when dealing with personal health information. CS3 notes that it is not necessary to obtain consent to process your health data when the data is used exclusively for medical diagnosis and health-care provision. However, the processing of personal data for marketing purposes requires the explicit consent of the patient. Furthermore, in all four organizations, mechanisms have been established to notify data breaches to the supervisory authority within 72 h.

*Summary of findings and main implications*

Table 3 aims to summarize the main findings identified distributed by software engineering practices. For each finding, the respective implications on the practical dimension of organizations to meet the challenges posed by GDPR are identified.

## Conclusions

With the implementation of the GDPR in May 2018, companies have felt an impact on their business strategies and adopted new procedures to collect, store and protect citizens' data. Software engineering companies have been one of the most affected sectors, as the application of GDPR has an impact not only on their internal processes but also on the various phases of the software development model.

Through four case studies with software engineering companies, it was identified that all phases of software development have changed, particularly in the early stages of the development of a new project. It is evident that in the early stages of project design, strategic

| Software engineering practices | Findings | Implications |
|---|---|---|
| Collection of requirements | – Data protection by default and design should be implemented continuously and from an early stage | – Integration of these practices in the DevOps process |
| | – Data processing activity and proportionality | – Minimization of the period of storing the information |
| | – Record activities performed on data | – Implement a data mapping taxonomy |
| | – Identify invasive data processes | – Implement a DPIA |
| | – Involvement and profile of the DPO | – DPO should have IT and law skills |
| Planning of activities | – Traceability of the information | – Adoption of IT tools to security and tracking access |
| | – Declaration of consent should be simple and clear | – Pre-validity options are not valid |
| | – Data security should be documented | – Ensure reliability, integrity and availability of access to data |
| Analysis and modeling | – High relevance of non-functional requirements | – GDPR should be considered a non-functional requirement |
| | – Effects on UX | – Evaluate the UX when applying GDPR compliance |
| | – Pseudonymization | – Encryption of data |
| | – Clarify the need of personal data | – Minimize data collection processes and guarantee their validity |
| | – Interoperability is a key factor in GDPR | – Adoption of open standards |
| Technical development | – Adaptation of internal processes and systems | – IT professions should be aware of the GDPR regime |
| | – High turnover rate of IT teams | – Need to always offer constant training programs |
| Testing | – Real data must be masked | – Adoption of automated tools for generating synthetic data |
| | – Delete data test when not needed | – Potential increase of testing phase without automatization |
| Deployment and delivery | – Increase awareness of information security and privacy | – Adoption of PETs |
| | – Difficulties in erasing data from public domains | – Increase coordination mechanisms with external parties |
| | – Identify and notify data breaches | – Implement an incident management policy |

**Table 3.**
Findings and their implication on software engineering practices

decisions must be taken regarding the volume and type of personal data requested and stored by applications. Consequently, the data owners' protection mechanisms should be carried out in the early stages of collecting the project requirements, in which the data protection by design and default paradigm should be followed. It was also discussed in this study how this paradigm can be incorporated into DevOps practices. Equally important is to consider data minimization and pseudonymization techniques using encryption and decryption methods. The DPO's involvement in these early stages of capturing the requirements of a project was also considered fundamental to identify the security and privacy risks that are inherent to each project.

In the early stages of a project, it is also necessary to define procedures for handling personal data and establish a plan to obtain the consent of the data owners. The

identification of non-functional requirements concerning documentation, security and privacy also gains greater relevance. From a technical point of view, new challenges arise such as the adoption of open standards and the creation of a user-friendly design for the user to manage his/her personal data. The UX should include not only access to the functional requirements of the application but also access to the users' personal data. This process gives more control over his/her data (e.g. approval, portability, right of rectification and right to forget) and more transparency on the purposes of collection and the storage period. However, the challenge lies in incorporating the obligations imposed by the GDPR with a smooth and pleasant UX that helps the company gain more credibility among its users. Already in the final phase of software development, changes in test procedures also arise, being important the adoption of synthesized data and the removal of test data when they are no longer needed. Also, in the deployment and delivery phase, some challenges arise related to the necessary mechanisms for data erasure, data audit and data breach notification. Furthermore, the adoption of PETs can be a supporting tool to streamline and simplify the process of privacy management and data protection. Finally, the technical development phase is the one that undergoes the least changes, but in which the importance of training software developers for GDPR rules and practices stands out.

This study addresses an emergent and still unexplored topic, as most studies on the impact of GDPR do not explicitly look for its effects on software development practices. This study takes a predominantly practical approach by identifying the challenges that are posed to organizations in each phase of software engineering practices, such as requirements gathering, activity planning, analysis and modeling, technical development, testing and deployment and delivery. This study is innovative in the conducted practical approach and offers mainly hands-on contributions, which is useful for software engineering companies to better prepare for the GDPR challenges they are facing. General guidance on the impact of GDPR on the software engineering process is complemented with examples of practices implemented by micro and SME software companies. This work has above all two important limitations that become relevant to mention. First, only four case studies involving two SMEs and one micro company were considered. In this sense and as future work, it would be important to seek to categorize the challenges posed according to the size of the companies. Additionally, the involvement of large companies in this process would be relevant. Another limitation of this study is that only the waterfall development methodology proposed by Sommerville was followed. Although this methodology continues to be used by most software companies, many of them in several projects are also beginning to strongly adopt agile methodologies. Agile methodologies processes tend to be less formal and the capture of requirements is evolutionary and gradual. In this sense, it will also be desirable to explore how the challenges identified in this study can be migrated to agile methodologies like Scrum or Kanban, particularly in capturing information security and privacy requirements.

## References

Almeida, F. and Monteiro, J.A. (2017), "Approaches and principles for UX web experiences: a case study approach", *International Journal of Information Technology and Web Engineering (Engineering)*, Vol. 12 No. 2, pp. 49-65.

Anirudh, M.K. and Massillamani, M.R. (2019), "Efficient cryptographic encryption techniques for data privacy preservation", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol. 8 No. 78, pp. 364-367.

Barrett, C. (2019), "What does GDPR mean for UX?", available at: https://uxdesign.cc/what-does-gdpr-mean-for-ux-9b5ecbc5 (accessed 12 March 2020).

Baruh, L., Secinti, E. and Cemalcilar, Z. (2017), "Online privacy concerns and privacy management: a Meta-analytical review", *Journal of Communication*, Vol. 67 No. 1, pp. 26-53.

Bendovschi, A. (2015), "Cyber-Attacks – trends, patterns and security countermeasures", *Procedia Economics and Finance*, Vol. 28, pp. 24-31.

Biscoe, C. (2017), "Data mapping: where to start for GDPR compliance", available at: www.itgovernance.co.uk/blog/data-mapping-where-to-start-for-gdpr-compliance (accessed 22 March 2020).

Bluestone, D. (2021), "State of GDPR in 2021: cookie consent for designers and developers", available at: www.smashingmagazine.com/2021/03/state-gdpr-2021-cookie-consent-designers-developers/ (accessed 6 March 2021).

Brodin, M. (2019), "A framework for GDPR compliance for small- and medium-sized enterprises", *European Journal for Security Research*, Vol. 4 No. 2, pp. 243-264.

Bygrave, L.A. (2017), "Data protection by design and by default: deciphering the EU's legislative requirements", *Oslo Law Review*, Vol. 1 No. 2, pp. 105-120.

Cascio, W.F. and Montealegre, R. (2016), "How technology is changing work and organizations", *Annual Review of Organizational Psychology and Organizational Behavior*, Vol. 3 No. 1, pp. 349-375.

Cavoukian, A. (2006), "Privacy by design: the 7 foundational principles", available at: https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf (accessed 5 July 2020).

Chen, Y., Ramamurthy, K. and Wen, K. (2015), "Impacts of comprehensive information security programs on information security culture", *Journal of Computer Information Systems*, Vol. 55 No. 3, pp. 11-19.

Clearwater, A. and Philbrook, B. (2018), "Privacy by design and GDPR: putting policy into practice", available at: www.cpomagazine.com/data-privacy/privacy-by-design-and-gdpr-putting-policy-into-practice/ (accessed 5 July 2020).

Clifford, D., Graef, I. and Valcke, P. (2019), "Pre-formulated declarations of data subject consent – citizen-consumer empowerment and the alignment of data, consumer and competition law protections", *German Law Journal*, Vol. 20 No. 05, pp. 679-721.

Crutzen, R., Peter, G.Y. and Mondschein, C. (2019), "Why and how we should care about the general data protection regulation", *Psychology and Health*, Vol. 34 No. 11, pp. 1347-1357.

Dove, E.S. (2018), "The EU general data protection regulation: Implications for international scientific research in the digital era", *Journal of Law, Medicine and Ethics*, Vol. 46 No. 4, pp. 1013-1030.

Duan, F. (2017), "The universal declaration of human rights and the modern history of human rights", available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3066882 (accessed 15 March 2020).

Esteve, A. (2017), "The business of personal data: Google, Facebook, and privacy issues in the EU and the USA", *International Data Privacy Law*, Vol. 7 No. 1, pp. 36-47.

Freitas, F., Ribeiro, J., Brandão, C., Azevedo de Almeida, C., Neri de Souza, F. and Costa, A.P. (2019), "How do We like to learn qualitative data analysis software?", *The Qualitative Report*, Vol. 24 No. 13, pp. 88-106.

Groen, M. (2019), "The anatomy of a non-intrusive, GDPR-compliant cookie message", available at: https://uxdesign.cc/the-least-obtrusive-and-gdpr-compliant-cookie-message-5df8b82fde8e (accessed 6 March 2021).

Guerriero, M., Tamburri, D.A., Ridene, Y., Marconi, F., Bersani, M. and Artac, M. (2017), "Towards DevOps for privacy-by-design in Data-Intensive applications: a research roadmap", *In Proceedings of the 8th ACM/SPEC on International Conference on Performance Engineering Companion, L'Aquila, Italy*, pp. 139-144.

Guinchard, A. (2018), "Taking proportionality seriously: the use of contextual integrity for a more informed and transparent analysis in EU data protection law", *European Law Journal*, Vol. 24 No. 6, pp. 434-457.

HIPAA Journal (2018), "GDPR: What is the role of the data protection officer?", available at: https://www.hipaajournal.com/gdpr-role-of-the-data-protection-officer/ (accessed 15 March 2020).

Hoofnagle, C.J., van der Sloot, B. and Borgesius, F.Z. (2019), "The european union general data protection regulation: what it is and what it means", *Information and Communications Technology Law*, Vol. 28 No. 1, pp. 65-98.

KPMG (2017), "The impact of the general data protection regulation in Portugal", available at: https://home.kpmg/pt/en/home/insights/2017/04/impact-of-gdpr.html (accessed 22 March 2020).

Li, Y. (2011), "Empirical studies on online information privacy concerns: literature review and an integrative framework", *Communications of the Association for Information Systems*, Vol. 28, pp. 453-496.

Li, H., Yu, L. and He, W. (2019), "The impact of GDPR on global technology development", *Journal of Global Information Technology Management*, Vol. 22 No. 1, pp. 1-6.

Macenaite, M. and Kosta, E. (2017), "Consent for processing children's personal data in the EU: following in US footsteps?", *Information and Communications Technology Law*, Vol. 26 No. 2, pp. 146-197.

Marinina, M. (2019), "AI, ML, and data analytics in the age of privacy regulations", available at: https://towardsdatascience.com/ai-ml-and-data-analytics-in-the-age-of-privacy-regulations-2b79447d5239 (accessed 22 March 2020).

Maynard-Atem, L. (2019), "The data series – solving the data privacy problem using synthetic data", *Impact*, Vol. 2019 No. 2, pp. 11-13.

Monreale, A., Rinzivillo, S., Pratesi, F., Giannotti, F. and Pedreschi, D. (2014), "Privacy-by-design in big data analytics and social mining", *EPJ Data Science*, Vol. 3 No. 1, pp. 1-26.

Mulligan, D.K., Koopman, C. and Doty, N. (2016), "Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy", *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, Vol. 374 No. 2083, id. 20160118.

Nasir, A., Arsah, R.A., Hamid, M.R. and Fahmy, S. (2019), "An analysis on the dimensions of information security culture concept: a review", *Journal of Information Security and Applications*, Vol. 44, pp. 12-22.

Neumann, G.K., Grace, P., Burns, D. and Surridge, M. (2019), "Pseudonymization risk analysis in distributed systems", *Journal of Internet Services and Applications*, Vol. 10 No. 1, pp. 1-16.

OECD (2020), "Employees by business size (indicator)", available at: https://data.oecd.org/entrepreneur/employees-by-business-size.htm#indicator-chart (accessed 23 (March 2020).

OPCC (2017), "Privacy enhancing technologies – a review of tools and techniques. Office of the privacy commissioner of Canada", available at: www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2017/pet_201711 (accessed 22 March 2020).

Politou, E., Alepis, E. and Patsakis, C. (2018), "Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions", *Journal of Cybersecurity*, Vol. 4 No. 1, pp. 1-20.

Poritskiy, N., Oliveira, F. and Almeida, F. (2019), "The benefits and challenges of general data protection regulation for the information technology sector", digital policy", *Digital Policy, Regulation and Governance*, Vol. 21 No. 5, pp. 510-524.

Queirós, A., Faria, D. and Almeida, F. (2017), "Strengths and limitation of qualitative and quantitative research methods", *European Journal of Education Studies*, Vol. 3 No. 9, pp. 369-387.

Reynolds, C.J. and Wyatt, J.C. (2011), "Open source, open standards, and health care information systems", *Journal of Medical Internet Research*, Vol. 13 No. 1, id. e24.

Romero, S. and Heredero, C. (2017), "Contribution of privacy by design (of the processes)", *Harvard Deusto Business Research*, Vol. VI No. 3, pp. 176-191.

Schartum, D.W. (2016), "Making privacy by design operative", *International Journal of Law and Information Technology*, Vol. 24 No. 2, pp. 151-175.

Sommerville, I. (2015), *Software Engineering*, Pearson, London.

Starchon, P. and Pikulik, T. (2019), "GDPR principles in data protection encourage pseudonymization through most popular and full-personalized devices - mobile phones", *Procedia Computer Science*, Vol. 151, pp. 303-312.

Taylor, S. (2020), "2020 Cybersecurity statistics, threats, and mitigation options", available at: https://restoreprivacy.com/cyber-security-statistics-2020/ (accessed 23 March 2020).

Teixeira, G.A., da Silva, M. and Pereira, R. (2019), "The critical success factors of GDPR implementation: a systematic literature review", digital policy", *Digital Policy, Regulation and Governance*, Vol. 21 No. 4, pp. 402-418.

Tikkinen-Piri, C., Rohunen, A. and Markkula, J. (2018), "EU general data protection regulation: changes and implications for personal data collecting companies", *Computer Law and Security Review*, Vol. 34 No. 1, pp. 134-153.

Titus (2018), "GDPR makes employee data security education essential", available at: https://titus.com/blog/compliance-regulation/gdpr-makes-employee-data-security-education-essential (accessed 22 March 2020).

Truong, N.B., Sun, K., Lee, G.M. and Guo, Y. (2020), "GDPR-compliant personal data management: a Blockchain-Based solution", *IEEE Transactions on Information Forensics and Security*, Vol. 15, pp. 1746-1761.

Van Dijk, N., Rommetveit, K. and Raab, C. (2018), "Right engineering? The redesign of privacy and personal data protection", international review of law", *Computers and Technology*, Vol. 32 Nos 2/3, pp. 230-256.

Wachter, S. (2019), "Data protection in the age of big data", *Nature Electronics*, Vol. 2 No. 1, pp. 6-7.

Wiesemborski, M. (2019), "How to design with privacy in mind | on privacy by design", available at: www.getrevue.co/profile/martinwiesemborski/issues/how-to-design-with-privacy-in-mind-on-privacy-by-design-178180 (accessed 5 July 2020).

Yin, R. (2017), *Case Study Research and Applications: Design and Methods*, SAGE Publications, Thousand Oaks, CA.

## Further reading

Creswell, J.W. and Poth, C.N. (2017), *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, SAGE Publications, Thousand Oaks, CA.

## Corresponding author

Fernando Almeida can be contacted at: almd@fe.up.pt