Data Security on the Ground: Investigating Technical and Legal Requirements under the GDPR

Maria Konstantinou*
Tina Marjanov*
m.konstantinou@student.vu.nl
t.marjanov@student.vu.nl
Vrije Universiteit Amsterdam
Amsterdam, Netherlands

Magdalena Jozwiak
m.e.jozwiak@vu.nl
Vrije Universiteit Amsterdam
Amsterdam, Netherlands

Dayana Spagnuelo
d.spagnuelo@vu.nl
Vrije Universiteit Amsterdam
Amsterdam, Netherlands

CCS CONCEPTS

• **Security and privacy → Privacy protections**; **Security services**.

KEYWORDS

GDPR, Article 32, security of processing, risk assessment, technical requirements, legal requirements

ACM Reference Format:
Maria Konstantinou, Tina Marjanov, Magdalena Jozwiak, and Dayana Spagnuelo. . Data Security on the Ground: Investigating Technical and Legal Requirements under the GDPR. In *womENcourage '21: ACM Celebration of Women in Computing, September 22–24, 2021, Virtual - Europe and Beyond*. ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

While the General Data Protection Regulation (GDPR) has been in force in the EU since May 2018, there is still much uncertainty on how to meet its demands in practice. For instance, in Article 32, the regulation defines that the data controller "shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk". The GDPR gives some indication on the aspects that should drive the decision on appropriate measures, but it admits multiple interpretations. Thus, when reading the regulation's demands, one question resonates: How to devise and put in practice technical measures suitable to guarantee such technical and legal demands from Article 32?

So far, researchers have tried to determine requirements for compliance through the use of established industry standards, stakeholder interviews and analyzes of state-of-the art technology [2, 4]. Such works tend to look at specific technologies or discuss only the highest level of compliance, often overlooking the common cases where such requirements are not strictly necessary.

If, on one side, we lack concrete guidelines on how to comply with GDPR's demands, on the other, information on what is not compliant is already available: hundreds of yet under-explored fines[1] have been imposed on the basis of violation of Article 32. These fines contain a wealth of information about the legal interpretations of the GDPR in practice, frequent violations, and suggestions by the various Data Protection Authorities (DPAs). In this project, we extract such information from the fines and analyze it from both technical and legal perspective. This allows us to get a realistic view on the common problems as well as the appropriate and practical solutions.

2 METHODOLOGY

The core of this project is the analysis of cases of infringement of Article 32 GDPR. In order to capture the requirements set forth in this provision, we identified the most important technical and legal aspects of the scrutinized cases. We did so based on the literature review and preliminary analysis of the sample cases. This information is laid out in a joint ontology, later used as a blueprint for the case annotation and analysis. The ontology captures basic concepts of a case, as well as technical and legal details that allow for deeper analysis. A subset of the most relevant concepts for the substantive analysis of cases is shown in Table 1.

Table 1: Ontology categories with possible values

Category	Possible values
GENERAL	
Decision, Country & DPA, Date, Fine, Other GDPR articles	
TECHNICAL	
Stage of processing	Collection/Storage/Processing/Destruction
Origin of threat	Internal/External
Maliciousness	Yes/No
Data type and format	Digital/Analog, Text, Pictures, Video
Mistake type	Human, Organizational, Technical
Requirements broken	Access control/Confidentiality/Integrity/Availability/Test & audit
LEGAL	
Data breach	Yes/No
Type of data breach	Unauthorized access/Unlawful processing
GDPR infringement	Inherent risk in large datasets/Sensitive data
Cause of infringement	Inadequate implementation of security measure/Faulty code
LEGAL - RISK FACTORS	
Scope of processing	Increased data quantity/State organizations holding large datasets
Nature of data	Financial/Health/Educational
Type of data subject	Students/Patients
Type of data controller	State/Private organization, Large/Small
LEGAL - HARM	
Likelihood	Low/Medium/High
Severity	Low/Medium/High
Type of harm	Material: Identity theft/Financial loss Moral: Emotional distress/Chilling effect
Right/freedom affected	Privacy/Expression
LEGAL - TECHNICAL & ORGANIZATIONAL MEASURES	
Operational readiness	Staff training/ISO implementation
Remedies post factum	Swift notification/Operational remedies

The sample of cases we analyze includes English translations of decisions issued in different EU countries and the UK –available on GDPRHub. In choosing which cases to include into the sample, we considered cases representing a wide range of countries, data controllers, violation types and case complexities. We also consider case availability and level of detail given. Since this is an ongoing

*Both authors contributed equally to this research.
[1] As of 30.6.2021, according to PrivacyAffairs

project, the present paper is based on a representative subset of 20 decisions we have fully analyzed. For each case, we first annotated the defining characteristics according to the ontology, using an online tool Hypothes.is. Then we qualitatively and quantitatively analyzed the collected dataset and documented observed patterns, together with preliminary guidelines.

3 RESULTS AND DISCUSSION

We observe a wide variety in the nature of the breaches. Depending on the type of data controller, type and scale of data, nature of mistake, and other technical characteristics we outline four risk groups and corresponding suggestions and appropriate data handling practices:

(1) **Coordinated high tech attack** (*e.g.*, targeted malware, cross-site scripting): Data controller is typically a bigger private or public organization with large amounts of data involved in a breach. The resulting fines are typically large and can reach several million Euros. Such organizations should aim to comply with the highest industry standards (ISO/IEC 27000 family), perform regular auditing and testing, and employ dedicated staff to ensure security.
(2) **In breach of GDPR, but without an incident** (*e.g.*, insufficient access control, insufficient auditing and protocols, too broad authorization): All types of data controllers are involved, most commonly medium or large organizations. The fines are often comparable to risk group (1). Given the organizational nature of the breach, it can be mitigated by performing a threat analysis to identify the weak points of the system and address them. Depending on the resources, suggestions from risk groups (3) or (1) also apply.
(3) **Moderate breach caused by human oversight or technical mistakes** (*e.g.*, database leaks, unauthorized sending of data, unencrypted data on websites): Data controller is typically a mid-scale organization, with minimal or no technical staff. The breach can involve one up to several hundred people, which is evident in fines ranging from several thousand to hundreds of thousands of Euros. The mistake is often caused by a misconfiguration, system reset or release of test data. Such cases show lack of protocols and established processes for handling the systems.
(4) **Low tech, human mistakes** (*e.g.*, wrong email attachments, unattended computers, improper disposal, mass emails without BCC *etc.*): Data controller is usually a small organization or an individual. The breach commonly affects a single person, or reveals only non-sensitive personal data, such as an email address. Fines are usually accordingly low. The implementation of advanced data handling systems is often unnecessary or impossible. Instead, the controller should try to improve awareness (what data is sensitive and who can access it) and good data practices (password protect machines/files, proper storage, safe electronic communication).

In addition to the risk groups, we identify some common danger points, with higher likelihood of an incident, and consequently a breach of GDPR. Such danger points include system reset, restore or restart, platform and version migrations, and outsourcing or sharing work on parts of a system.

In line with the technical perspective and the classification of the findings into risk groups, the DPAs apply a risk-based approach into the legal argumentation of the decisions and the imposition of the fines. Under this approach, a data processing activity is undertaken only if the risk is not too high or if it can be sufficiently managed, in compliance with the GDPR. Accordingly, the greater the risk and the imminent harm, the greater the GDPR fine. Such a scalable risk-based approach to DPA decision-making allows for the imposition of fines even when there is no data breach, but merely an infringement of the GDPR that could potentially (but not surely) raise risks and inflict harms to the data subject(s).

Absent from the analyzed sample of GDPR fines is an in-depth examination of the ways in which such risks could lead to specific material or moral harms, as well as how such harms (would) compromise fundamental rights. Thus, the DPAs apply the risk-based approach rather than the rights-based approach to data protection in the EU, which would be more in line with safeguarding the essence of data protection as a fundamental right, placing it at the center of GDPR compliance assessment [1, 3]. Simultaneously, it could also help avoid a scalable risk analysis, inefficient organizational measures and, consequently, unconsidered, long-term harms.

Methodology-wise, most DPAs adequately investigate the factual and substantive elements of each Article 32 case, often proposing technical and organizational mitigation measures. However, differences are observed in the degree to which DPAs substantively analyze each case and the extent to which they engage with the fundamental rights framework in their legal analysis.

4 FUTURE RESEARCH

As this is work in progress, the sample size as well as the conclusions are not yet finalized. We aim to expand the sample size and ideally pick up on more subtle properties of the cases. The current annotation process is manual and rather slow and could, as such, benefit from automation in the form of text analysis through AI. The annotation and other raw data of this project will be made available to future researchers through the use of GDPRHub and public annotations.

ACKNOWLEDGMENTS

This research was conducted with the support of the Network Institute of Vrije Universiteit Amsterdam.

REFERENCES

[1] Maja Brkan. 2019. The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning. *German Law Journal* 20, 6 (2019), 864–883.
[2] Vasiliki Diamantopoulou, Aggeliki Tsohou, and Maria Karyda. 2020. From ISO/IEC27001: 2013 and ISO/IEC27002: 2013 to GDPR compliance controls. *Information & Computer Security* 28, 4 (2020), 645–662.
[3] Raphaël Gellert. 2020. *The Risk-Based Approach to Data Protection.* Oxford University Press, Oxford.
[4] Sean Sirur, Jason RC Nurse, and Helena Webb. 2018. Are we there yet? Understanding the challenges faced in complying with the General Data Protection Regulation (GDPR). In *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security.* ACM, Seoul, 88–95.