Informatics

Datenschutz mit oder gegen Blockchain-Technologie

Eine Analytische Reflexion und Diskussion der Ausübung der sich aus der DSGVO ergebenden Rechte zur Verbesserung der Privatsphäre im öffentlichen digitalen Raum beim Einsatz der Blockchain-Technologie

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Sozial- und Wirtschaftswissenschaften

eingereicht von

Dipl.-Ing. Dominik Schmelz, BSc
Matrikelnummer 00727975

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Thomas Grechenig

Diese Dissertation haben begutachtet:

Christian M. Piska

Wolfgang Slany

Wien, 24. Jänner 2023

Dominik Schmelz

TU WIEN Informatics

Data Protection via or versus Blockchain Technology

An Analytic Reflection and Discussion of the Exercise of Rights Arising from the GDPR to Improve Privacy in the Public Digital Space when using Blockchain Technology

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Sozial- und Wirtschaftswissenschaften

by

Dipl.-Ing. Dominik Schmelz, BSc
Registration Number 00727975

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Thomas Grechenig

The dissertation has been reviewed by:

Christian M. Piska	Wolfgang Slany

Vienna, 24th January, 2023

Dominik Schmelz

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Dipl.-Ing. Dominik Schmelz, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 24. Jänner 2023

Dominik Schmelz

Danksagung

Ich möchte mich bei allen Kollegen, Freunden und Familienmitgliedern bedanken, die mich bei der Erstellung dieser Dissertation unterstützt haben. Besonderer Dank geht an meine Frau, Lieselotte. Sie hat einen Teil unserer kostbaren Zeit geschenkt, damit ich diese Dissertation schreiben kann. Auf der technischen Seite möchte ich mich bei meinen Kollegen und Freunden Phillip Niemeier, Karl Pinter und Andreas Kern bedanken, die mich bei der Findung von Problemen und deren Lösung unterstützt haben. Nicht zuletzt möchte ich mich bei meinem Betreuer Thomas Grechenig für seinen Rat und seine Hilfe während des gesamten Prozesses der Dissertation bedanken.

Acknowledgements

I want to thank all my colleagues, friends and family members who supported me in writing this dissertation. Special thanks go to my wife. She has given a portion of our precious time so that I can write this dissertation. On the technical side, I would like to thank my colleagues and friends, Phillip Niemeier, Karl Pinter, and Andreas Kern, who supported me in finding issues and solutions. Last but not least, I would like to thank my supervisor Thomas Grechenig for his advice and help during the whole process of the dissertation.

Kurzfassung

Der Datenschutz ist in den vergangenen Jahren durch Fortschritte bei der Datenerfassung und -auswertung, Datenschutzverletzungen und strengere Rechtsvorschriften, wie die Datenschutzgrundverordnung (DSGVO) der Europäischen Union, in den Fokus der Öffentlichkeit gerückt. Die Blockchain hat nach ihrer erfolgreichen Anwendung beim Austausch von Vermögenswerten an öffentlichem Interesse gewonnen. Ihre Offenheit und Unveränderlichkeit haben Bedenken hinsichtlich des Datenschutzes und der langfristigen Auswirkungen auf das persönliche Dateneigentum geweckt, aber auch neue Ansätze zur Rückgewinnung individueller Rechte in einem derzeit zentralisierten Umfeld von Autorität und Kontrolle ermöglicht. Diese Dissertation widmet sich einer strukturierten technisch-rechtlichen Analyse zur Nutzung von Blockchains und deren Auswirkungen auf die Einhaltung der DSGVO, einer Diskussion über die Auswirkungen von Lösungen zur Minderung von Blockchain-Datenschutzrisiken und wie Blockchain-Technologie in Anwendungen zur Unterstützung des Datenschutzes eingesetzt werden kann, um die Rechte betroffener Personen gemäß der DSGVO einzufordern.

Diese Arbeit beinhaltet eine qualitative risikobasierte Analyse von Datenschutzproblemen mit Blockchains sowie eine Analyse von Lösungen, Risikobehandlungen und resultierenden Restrisiken. Darüber hinaus werden zwei wichtige DSGVO-Prozesse modelliert und potenziell problematische Prozessaktivitäten durch die Erleichterung quantitativer Forschung identifiziert. Diese werden verwendet, um Kriterien für unterstützende Anwendungen mit Hilfe der Delphi-Methode zu identifizieren. Diese Kriterien werden strukturiert und verdichtet, und es wird ein Kriterienkatalog generiert, um Anwendungen hinsichtlich ihrer Unterstützung zu bewerten. Zwei Fallstudien werden mit Hilfe einer Expertenevaluation auf die Kriterienfragen angewendet. Ergebnisse sind eine Methode zur risikobasierten Bewertung von Technologien im Bereich des Datenschutzes, eine Technik zur Modellierung von Prozessen in Gesetzen und zur Bewertung von Technologien hinsichtlich ihrer Unterstützung in diesen, die zeigen, wie Blockchain-Technologie bestimmte datenschutzrelevante Aktivitäten unterstützen kann.

Die Ergebnisse zeigen, dass einige aktuelle Lösungen, einen geringen Einfluss auf die Risiken haben; öffentliche Blockchains können in Anwendungen verwendet werden, um bestimmte Aktivitäten der GDPR-Prozesse der betroffenen Personen zu unterstützen. Die Ergebnisse ermöglichen es Forschern, Technologen und politischen Entscheidungsträgern, Technologien zu bewerten und ihre Entscheidungen auf ein wissenschaftliches Modell zu stützen. Weitere Forschungsarbeiten können die Risikoanalyse-Vorlage zur Bewertung von Technologien zur Verbesserung des Datenschutzes für Blockchain-Technologien nutzen, um das Risiko weiter zu verringern.

Abstract

Data protection came into public focus in past years through advances in data collection and evaluation, data breaches and stricter legislation such as the European Union's General Data Protection Regulation (GDPR).

Distributed ledger technologies such as blockchain have attracted public interest following their successful application in asset exchange. Their openness and immutability raised new concerns about data privacy and its long-term impact on individual data ownership, but also enabled new approaches to reclaiming individual rights in a currently centralised environment of authority and control. This dissertation addresses the need for a structured techno-legal analysis on the matters of uses of public and private blockchains and their impact on GDPR compliance.

This work contains a qualitative risk-based analysis of privacy issues with private and public blockchains, and an analysis of solutions, risk treatments and resulting residual risk. Furthermore, two important GDPR processes are modelled, and potentially problematic process activities are identified by facilitating quantitative research with a survey. These process models are also used to identify criteria for supporting applications using the Delphi method. These criteria are structured and condensed with the help of thematic maps, and criteria questions are generated to assess applications regarding their supporting features. Two case studies are applied against the criteria questions with the help of an expert evaluation.

The results are a method for a risk-based assessment of technologies regarding data protection, a technique used to model processes of laws and assess technologies regarding their support in these processes that show how blockchain technology can support certain data subject related data protection activities.

The results show that some solutions referred to in current literature have a small impact on the risks; public blockchains can be used in applications to impact certain activities of data subjects' GDPR processes. The outcomes allow researchers, technologists and policymakers to assess technologies and base their decisions on a scientific model. Further research can use the risk-analysis template for assessing privacy-enhancing technologies for blockchain technologies to further reduce risk.

Contents

CHAPTER 1

Introduction

Advances in computer science have raised concerns about technology-related privacy and protection thereof. Countries around the world are heading in similar directions to re-balance the rights between people, wanting to protect their privacy and businesses and governments wanting to profit from the collection and use of private data [UNC16]. In May 2018, the European General Data Protection Regulation (GDPR) [Cou16a] came into effect. Companies offering goods or services to European citizens (so-called data subjects) have been obligated to implement data protection measurements when dealing with private data. These include rights for the data subjects, the need for a legal basis for personal data processing, documentation of processing operations and technical obligations. The United States of America and other countries around the world implemented similar legislation in order to regulate the legal collection and use of personal data.

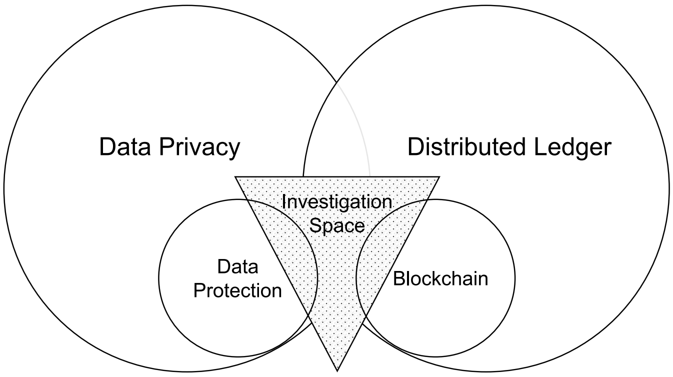


Figure 1.1: Investigation Space

Developments in Distributed Ledger Technology, most famously blockchain, have enabled new ways of sharing data without a central authority controlling the information flows

[CDP18]. The lack of central responsibility, the immutability, and the public accessibility of published data make blockchain complicated to be used in a data-protected way [SFN$^+$18]. Besides raising privacy concerns, blockchain could controversially be used to solve privacy issues.

Figure 1.1 shows the investigation space between data privacy and distributed ledger, especially blockchain and data protection. Based on the definitions of privacy and, accordingly, data protection by the GDPR, this dissertation investigates the space between data protection and blockchain by exploring, on the one hand, the causes and roots and potential solutions to data protection problems with blockchain technology and, on the other hand, dimensions and criteria for practical, data-protective solutions for data subjects based on blockchain technology to help them exercise their rights to data protection and privacy granted by the GDPR and thus ultimately improve data-protected public environments (see Figure 1.1).

Privacy issues of public blockchains have been discussed extensively in several publications, including one by the author in 2017. Some solutions to those problems already exist in the technical-legal literature. However, there are some controversies with those solutions. For example, encrypting publicly published data and then deleting the key is mentioned as a method to implement the right to be forgotten. These methods need to be summarised and discussed in a structured way.

Data subjects enjoy specific rights under the GDPR against the controller. Exercising these rights can be difficult for the data subject. Current procedures may even further endanger the privacy of the data subject. For example, data controllers are asked to send scans of IDs by mail in order to identify the data subject, which further burdens the data subject. Problems in the exercise of data subjects' rights must therefore be quantified and solved.

The current situation leads to the privacy of data subjects being endangered in both of these areas. Therefore, it is important to scientifically investigate and improve these areas in a targeted and structured manner.

The research questions are divided into two areas. Question 1 asks for privacy issues with blockchain technology, whereas Question 2 asks for solutions for defending data subjects' rights using blockchain technology.

Question 1.1: How does the use of a public or private blockchain change GDPR compliance in an application?

Question 1.2: What are feasible solutions for data protection issues on blockchains?

Question 2.1: How can blockchain technology be used in applications to support data protection according to the GDPR?

Question 2.2: How can blockchain-based applications to support data protection according to the GDPR help empower data subjects?

To answer the research questions, a structured approach is taken based on the Design Science Research Methodology (DSRM) [PTRC07]. This methodology is used since it is a multi-disciplinary problem that must be analysed from different perspectives. The DSRM approach allows using quantitative and qualitative research methods in a structured way to shed light on partial aspects. It also allows for social and legal analysis, which are both needed to find technical-legal solutions.

DSRM is structured in six process steps that are needed to scientifically find solutions to given problems [PTRC07]:

I Identify problem & motivate
II Define objectives of a solution
III Design & development
IV Demonstration
V Evaluation
VI Communication

First, the problem was identified, and the motivation to solve the problem needed to be found. In the research at hand, preliminary research was conducted and published that stated the problem with data protection and blockchain, but also the possibilities with the technology were outlined. Discussions with other researchers lead to a specific motivation and problem statement. The stated objective for this particular research was to make future blockchain development more data protected and to facilitate it for data protection. Blockchain development, in this case, means the development of blockchain technology itself but also the development of applications with the help of blockchain technology. The objectives of the studies and solutions were defined by designing a quantitative scale to test and measure the fulfilment of the objectives to support data protection in the means of the General Data Protection Regulation. This measurement used modelling of laws to assess applications designed in case studies to understand the environment and stakeholders and measure the impact of the applications. Following the objectives and problems, artifacts were created to enhance data protection with blockchain technology. These include a theoretic analysis to mitigate data protection issues and prototypical applications to facilitate blockchain technology in data subject supporting applications. These prototypes were tested with experiments in real-world scenarios and were feasible in case studies. The outcome of each case study was evaluated, and objectives were compared with observed behaviour. This was done with the created model and the assessment criteria defined with the help of the model. The problem identification and selected case studies were published and presented at international conferences to disseminate intermediate results.

The dissertation is structured as follows. First, the terms privacy and data protection are introduced and their terminology defined (Section 2.1). Then the current situation of data protection is presented, and a brief outlook on the future of data protection is given (Section 2.2). Then the legal basis for the following chapters is presented, namely the General Data Protection Regulation (Section 2.3).

In the following chapter, distributed ledgers (Section 3.1) and blockchain (Section 3.2) are introduced. Furthermore, it examines the technical details of blockchain technology that are necessary for the following discussion on privacy, namely data structures (Section 3.3) and protocols (Section 3.4). Finally, the chapter gives a brief overview of the current applications of blockchain technology (Section 3.5).

The following chapter focuses on the data-protected use of blockchains. It introduces and applies a risk-based assessment of personal data processing operations (Section 4.1) by evaluating private and public blockchains, exploring solutions to address the risks and assessing the residual risks (Section 4.2).

The following chapter presents an assessment process for technologies supporting legal processes by modelling processes in legal texts (Section 5.1), quantifying potential problems in the execution of these processes (Section 5.2), solution criteria are defined (Section 5.3), potential solutions to the problems are implemented (Section 5.4) and finally an evaluation of the solution is made (Section 5.5).

In the following chapter, use is made of this process by modelling two processes of the General Data Protection Regulation (Section 6.1), quantifying problems stakeholders have with these processes using quantitative research (Section 6.2), defining solution criteria using a Delphi panel (Section 6.3), presenting two practical case studies (Section 6.4) and the evaluation of the case studies with the help of expert interviews (Section 6.5). Finally, the results are reflected, and conclusions are drawn (Chapter 7).

CHAPTER 2

Data Protection

Data protection has been shifting into public focus in the last couple of years. This was due to incidents[1] and world-wide legal regulation (e.g. GDPR, CCPA, PDPA). Before that, it has been a field of low legal and technical restrictions. These allowed a rise in data collection and analysis applications, so-called big data applications [Sta19]. These summarise a set of technologies that aim to collect and analyse larger datasets in shorter- or even real-time.

Facing the possibility of breaking data protection laws and getting fined, companies started to implement GDPR and CCPA urgently before they became effective. Most levied GDPR fines were regarding the appropriate technical or organisational measures [Bar20b]. According to a survey [ct20], data protection regulators have imposed 114 million Euro in fines under the GDPR until January 2020. Many companies still struggle with implementing the requirements of the GDPR. A study in 2020 in Austria [Ruz20] showed that only 32% of companies have finished implementing their GDPR measures. Most of the companies asked struggled with deletion concepts and process implementations.

The following sections will give an overview of data privacy and data protection in general and a detailed, technical view of the EU's GDPR.

2.1 Privacy and Data Protection

Understanding the cause, use, and needs behind privacy is needed in order to be able to design mechanisms to protect it. Privacy is not just a technological phenomenon that was created because of technological advances. It became more important through the technological possibilities of collecting and sharing more information faster. Human

[1] Some of the best known: LinkedIn in 2021; Facebook in 2019; Equifax in 2017; Uber in 2016; Adobe in 2013 [Tun22]

beings use sharing of information as a technique of survival. Sharing of information is an important factor in being social. As human beings, we value socialisation and share information about ourselves with each other. The amount of personal information willing to share depends on the social relationship. One of the aspects of privacy is the control of what information is shared about oneself and limiting the access others have to one's personal information [Sol08].

Within this definition lies the assumption that the good being protected is information that we are not willing to share openly about ourselves. Information is a product of data and other information. It is data that has been brought into context, categorised, analysed mathematically or statistically, corrected, and condensed [DP98, p. 4]. GPS points of a person's smartphone are considered data. A route or location profile would be considered information. Resources in data protection tend to confuse data and information, including the GDPR, that defines 'personal data' as "information relating to an identified or identifiable natural person" [Cou16a, Art. 4].

Privacy has grown in significance in many fields. Definitions vary based on different contexts and aspects of society. The difference in most of the definitions is mainly due to the contexts of the environments from which the definitions exist. Privacy has social, psychological, technical, and legal aspects. Each field defines privacy from another perspective. They share the advocacy for personal space and freedom, though with different whereabouts and intentions.

The social aspect of privacy makes human beings cautious during social interactions by limiting the level of socialisation. Consequently, humans do not expose all their activities to each other and prefer to keep the information contextual. People value their privacy and appreciate that some areas of their lives are protected. Different circumstances in life make human beings desire that their personal information remain discrete and not accessible to anyone at any time. Social stigma [Gof63] may arise when information about oneself is disclosed, resulting in social discredit and loss of trust in society. Other social theories argue that privacy is less about controlling, stopping or reducing sharing of information but rather serves a social function [Wal18, p. 50]. Intimate situations may make us share more personal information about ourselves in order to build a special, intimate relationship [Ger78]. Gerstein even argues that an intimate relationship is not possible without privacy.

Psychological aspects regarding privacy focus on the subject disclosing personal information and the effects on itself. Studies in early child development show that privacy is an important factor in a child's early development [Wol78]. It has shown that children need privacy for developing their own solutions and ideas; a lack of privacy slows their development in some phases. Advancements and changes in society have made privacy invaluable to people, making privacy become essential in their lives. A rather interesting aspect of psychology and privacy is the 'privacy paradox' [Bar06]. It describes the dissonance between the intention of people and their actual behaviour regarding privacy [NHH07]. Chellappa and Sin [CS05] tried to pinpoint the behaviour by giving participants of an experiment the decision whether they wanted to customise their experience by providing personal information or staying more private. Their experiment showed that in the real-world decision, the use of ‘personalisation services’, and with it the disclosure of personal information, is almost twice as important as their concern for privacy. They also found that the perception of risk and trust influences their decision in regard to giving up privacy. The research suggests that the balance between people and businesses is actively influenced by the trustworthiness of businesses and the return people get in exchange for their information.

Legal Aspects of privacy deal with the definition of the intangible good that needs to be protected. Depending on the legislation, data or information about a person or a very small group of persons (households, families etc.) are regulated. The regulation usually defines the means of how, under what circumstances and by whom this intangible good is allowed or forbidden to be collected, stored, and used. Usually, it puts those involved in such activities under obligations, such as sharing information about activities regarding it with affected parties and authorities and protecting of the good. Countries that signed the The Universal Declaration of Human Rights must implement a privacy law. They signed that a human being has the right not to be “subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation” [Ass49, Article 12] and the right to be protected by a law against such interference.

Technological aspects of privacy focus on the developments of data collection, information retrieval and technical aspects of securing information from unauthorised access. It is, therefore, a collective term applied in various technical aspects regarding the access, control, storage, and use of personal information. Developments in technology allow an easier collection of data on personal information. Technology companies, such as Facebook and Twitter, rise in value because of their gathered information about their user. The so-called big data applications allow the extraction of information on large datasets that grow continuously. The marketing term ‘big data’ itself is less scientific but is surrounded by topics such as the advances in information technology, the rise of big technological companies, globalisation, global social networks, and the development of a platform economy that allows storage and processing of significant amounts of data [Flo14]. An invasion of privacy is, on the opposite, defined as processing without consent of the involved natural persons, either by governments, public or private organisations, or other individuals [MSP15].

Privacy and security are related. To understand their relationship and to draw conclusions based on these understandings, a structured, contextual elaboration of their relationship must be done (see Figure 2.1). ‘Security’ is defined as “freedom from, or resilience against, the potential harm caused by others” [Hol20]. Hollnagel adds the need for the detection of harm and the fast recovery from harm to the definition. ISO relates security to an asset that needs protection, such as information [ISO18]. ‘Cybersecurity’ is a real subset of security. The term cybersecurity is a less technical term. It is used in legislation, and therefore by politicians and rarely by practitioners [SBW17]. Schatz et al. try to narrow the term by analysis of its use and create a more comprehensive definition of

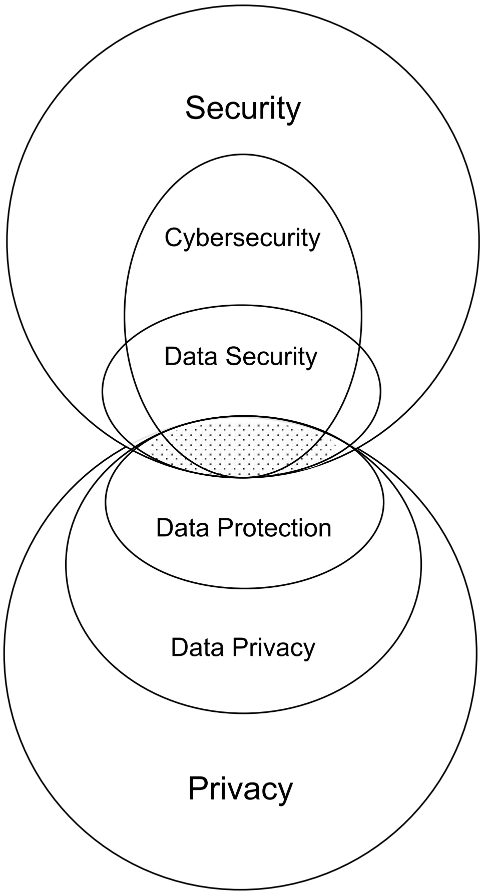


Figure 2.1: Relation of Security and Privacy

cybersecurity: "The approach and actions associated with security risk management processes followed by organisations and states to protect confidentiality, integrity and availability of data and assets used in cyber space" [SBW17].

'Data security' as "the science and study of methods of protecting data in computer and communication systems from unauthorised disclosure and modification" [Den82] is the technical implementation needed in computer systems needed for data protection. Data security strategies have evolved to ensure that only the authorised accesses with the consent, or legal basis, of the concerned individuals, can be done. Technological advancements also ensure that data recovery is possible after loss while also protecting data from compromise by all means possible while also facilitating data privacy [EERM15]. The purpose of data security is to control protective measures and access to data under different circumstances. It concerns both the operational backup of data to enable data recovery after data loss or compromise and the controlled and regulated backup of critical data. Data security is not a real subset of cybersecurity since it also encompasses analogue or offline data. Some argue that cybersecurity is, therefore, a superset of data security.

Using the definitions above, cybersecurity is neither a real subset nor a superset of data security. Information Security is often mixed with data security. Information Security sees information, rather than data, as an asset to be protected and is formally defined by the ISO as: “preservation of confidentiality, integrity and availability of information” [ISO18] in short CIA (see Section 2.3.3 for details on CIA). CIA does not apply in absolute terms but is risk management in relation to the intangible/material assets, in this case, information, worth protecting.

Data protection and data privacy are similar to privacy itself, defined by different fields and, therefore, different aspects. ‘Data protection’ though defined as “legal control over access to and use of data stored in computers” [Ste10] is usually used in context as the protection of personal data, not any data, and especially not only in computers but also analogue data and during transfer. It is more a technical term that encompasses management, including but not limited to access management and availability of data. Data privacy, sometimes wrongly used synonymously with information privacy, is a broad term that includes the technical and organisational protection of data, but according to Michael & Michael [Mic13] also all issues surrounding and especially also the relationship between,

- collection and dissemination of data,
- technology,
- the public expectation of privacy,
- legal issues,
- and political issues.

They show the wide field of data privacy by adding the aforementioned fields without naming them but rather their implications. The definition also specifically mentions the relationship between the aspects. Political discussion from all parties in power leads to legislation and legal issues. The public expectation of data privacy is more of a psychological and sociological issue, as mentioned above. It also encompasses differences in culture in different regions, such as China without any privacy culture or legislation or the United States with a strive for protecting their freedom of speech [Wan11].

Cybersecurity has evolved to become a conceptual debate that focuses on drawing a safe line between information technology and privacy. The borderline is characterised by laws that protect privacy by allowing them rights over their data by regulating how companies and governments use data [SS14]. Some regulators enforce the data protection laws in the complex environment and handle the challenges posed by new forms of technologies, models, systems, and services. The need for data protection is increasing in demand now more than ever due to the daily interaction with environments that generate and collect data on and from human behaviour. The framework of data security has a long way to go. There is a need for upcoming and existing regulatory regimes to formulate strategic methodologies to address the new data-intensive systems and frameworks [EERM15]. Ultimate data protection is inconceivable. The likely scenario is a world where there are fundamental legal and regulatory safeguards that have been implemented to provide much-needed governance to the national and global frameworks to avoid data exploitation.

The technological complexity and the complexity in governing how policies in governments and multinational corporations govern how information is collected, stored, and used are growing. Security is needed for the protection of privacy since data cannot be held private in an insecure environment. This involves much more than 'cybersecurity', but also buildings, organisations, governments, or any technical facilities and devices. This issue has been known for decades. Firnberg [Fir78] argued the need for security to achieve privacy in the late 70s. His risk-based approach can be found in many standard security standards these days. Security is also the ability to reinforce the available infrastructure while also considering existential regulations on privacy and confidentiality. Latest political discussions on anti-money laundering and the fight against terrorism deepened in the European Union [Com18]. The discussions on the topic of implementing backdoors in communication protocols for governmental law enforcement [Koo19] show the discrepancy between privacy, security and protection of the citizens by the government. The irony in this discussion lies in the fact that the same government that is responsible for enforcing standards for security to protect the privacy of their citizens has its own interests in insecure communication to violate their privacy. The discussion on privacy versus law enforcement is not dependent on the technology used. It has existed for a long time, at least since the 'secrecy of letters' or the 'privacy of correspondence' [R+09]. Trepete, in contrast, argues that privacy is increasingly a concern due to the increase of capacity in technology to breach the same [TTM+15]. The advances in information technology for sure made the amount of data, gathering of it and extraction of it easier, or even possible. They also made access to critical information easier, making it more important for the individual. It also increases the scope from which privacy can be intruded unconventionally.

Many technological achievements can be used for both good and evil. They might have been invented for the right reasons but can be used for unethical purposes, either. Technologies such as facial recognition are highly viable to maintain security, but in the wrong hands are potentially dangerous. Data analysis shows advances to de-anonymise information that was thought to be anonymous. [NS08] Internet of Things and smart mobile applications solve many issues but create data trails that might be used for analysis or surveillance [Ste14]. Health tracking increases the possibilities in diagnostics but creates the need for digital privacy in sectors that handle highly sensitive data which has not been recorded or accessible in this broad manner [ZMVOM18]. These advances generally create value for their users but also create a need for data protection. Data protection itself has become a complicated and sensitive issue involving the confidentiality of personal and critical information. The issue remains controversial in its importance and value, mainly due to the increasing development of new technologies and the lack of clarity, consensus and agreement on where the boundaries should be drawn in terms of laws, policies and ethics related to personal freedom.

Data protection only operates within a realm determined by the role of the individual in society. There reaches a point where the relevant authorities must invade the privacy of an individual due to security issues or to find incriminating evidence [KL14]. The regulations are, therefore, not a shield for criminals but rather a protection of the public. Each individual gains the right to protect itself against an involuntary invasion of privacy by others, including public authorities. Public authorities, in contrast, often enforce data protection and privacy. Privacy Enhancing Technologies (PETs) protect and preserve the confidentiality of the individuals by technology rather than legal protection. They preserve privacy by "eliminating or minimising personal data, thereby preventing unnecessary or unwanted processing of personal data, without losing the functionality of the information system" [VBBO03]. These technologies are often used to implement and comply with existing laws. However, national security experts often gather evidence by monitoring specific individuals' alarming activities and taking targeted action against criminals. [Ros14]. Cybersecurity is becoming a threat to national security, making it even more reason to have elite security forces that operate within the network. Although data protection and privacy are implemented, there are few loopholes, and criminals can rarely escape due to the readily available incriminating evidence of their activities online, and offline [KL14]. Moreover, individuals also play a significant role in maintaining their privacy and ensuring data security. When using the internet, it is the user's responsibility to ensure they access a safe website that cannot invade their privacy or data. Therefore, the latter means that privacy and data protection do not cover for lack of security and safety of the individual, which becomes a personal responsibility [EERM15]. Cybersecurity and privacy vary among different countries and in various private and public sectors. However, the regulations ensuring cyber-security and privacy harmonise and strengthen data protection while also ensuring effective law enforcement.

2.1.1 Reasons for Data Protection

Data protection and privacy play a crucial role in determining effective and democratic governance in which people enjoy their rights. Awareness of the right to privacy and data security has increased. However, few institutions and governments have the necessary legal and institutional infrastructure and processes to grant these rights to the public. Some countries around the world also lack essential regulatory and legal frameworks, while in other countries, the implementation and enforcement of these rights are inadequate [TPRM18]. The lack of data security and privacy frameworks sees technological innovations, policies, and practices unregulated and unchecked. Such have severe implications on individuals' rights and in the development of the affected economies and societies. The realisation of the importance of cybersecurity and privacy has encouraged innovation in data management [VBBO03]. Technology providers have introduced numerous advancements, such as encryption technologies that also hide data from criminals. The latter also makes the data that was once readily accessible to law enforcement unavailable. Law enforcement agencies seek the ability and power to bypass the available security protocols due to national and physical security concerns [Rau14].

Increased cybersecurity and national security threats such as terrorism are relatable in the modern world. Service providers are increasingly becoming compelled to re-engineer the products to allow bypassing of encryption protection to enable access to consumer data. However, these requests give the law enforcement agencies and, ultimately, the government powers that undermine the claim to protect the security and privacy of the citizens [DDFH+15]. Technology companies that leave backdoors to be used by law enforcement agencies and the government also make the workarounds target of cybercriminals, including hackers and hostile nations. Such vulnerabilities offer opportunities for data compromise and attack while also undermining the efforts of providers to secure consumer data. European Union Agency for Law Enforcement Cooperation (EUROPOL) stated that “a flaw in the encryption would not only enable access to law enforcement, but it would also make systems vulnerable to criminals. Therefore, most governments have now abandoned plans of backdoor-access schemes” [Rep19].

The need for consensus on data protection and privacy is that backdoors to allow access to consumer data will only create powerful vulnerabilities despite the struggles to address the existing challenges. The vulnerabilities could potentially affect governments and citizens as the cybercriminals could capture information from critical security infrastructures such as services and devices used by law enforcement and national security agencies [RR14]. Governments are increasing their efforts to improve responsibility regarding the security of consumer data, more so after breaches that exposed billions of customers' information to hackers by analysing occurred breaches and setting guidelines [Lew13]. The regulations should aim to ensure that government-mandated backdoors are not there as they can be used by hackers to compromise user data [Wai17]. Public and private institutions bear a fiduciary duty to protect consumer data, which ranges from personally identifiable information to intellectual property, national security secrets, and financial information. Data protection aims to ensure that companies can protect consumer data by retaining control of fundamental security decisions. Decisions relating to cybersecurity and privacy, more so the sweeping powers demanded by law enforcement agencies, should be reached based on a consensus [BR17]. The private sector and government should balance the interests of security and privacy. It is worrying that if data security and privacy are taken for granted, governments would sacrifice the ability of citizens and organisations to secure data and systems in search of vague security concerns.

A more unobvious reason for data protection is politics. Vogel argues that the European Union's GDPR is a strategy to push for European influence and agenda. He categorises them in strive for legitimating, creation of a global profile, and economic [Ros14]. Rossi argues against it by referring to several sources from outside the European Union has welcomed the GDPR. As a matter of fact the GDPR has influenced several laws including California's California Consumer Privacy Act (CCPA), China's Personal Information Protection Law of the People's Republic of China (PIPL), Barsil's Lei Geral de Proteçao de Dados (LGPD), and Thailand's Personal Data Protection Act (PDPA).

Data is becoming more valuable hence the need to protect it. The lack of enough data protection and privacy laws exposes the population to significant risks. The skills and opportunities for data retrieval are evolving, making identity thefts and other sabotage common and potentially dangerous situations. The lack of data security violates the fundamental freedom and rights of individuals related to the data. Data protection is conducted concerning the rights and freedoms of individuals. Lack of enough data security, such as for financial and personally identifiable data, exposes people to financial theft and fraud as well as life-threatening situations when health information is manipulated [Edw16]. The regulations ensure fair and consumer-friendly economic environments. The regulations have created an environment where economies and businesses thrive based on consumer power. Personal data regulations prevent the selling of personal data and give consumers control in the market. Leaking of personal data has, in the past, cost companies significant impairment to reputation and hefty penalties. All these scenarios emphasise the need to comply with existing data protection regulations and privacy. The regulations emphasise that the security of data comes from knowing what data is processed, on what grounds, and why it is being processed [BB16]. Companies are also tasked with ensuring that safety and security measures are in use during data protection. The advancements in technology have seen significant amounts of data stored digitally. Organisations are therefore tasked with protecting the information as a legal necessity as well as in protecting the organisation and economy. Different organisations store different critical pieces of information ranging from employee records to customer details. The information should be safeguarded as third parties can use the information for frauds such as phishing scams and identity theft. Identity theft inflicts damage to businesses, the victims of identity theft, the justice system, and society produce non-financial personal damage [NM05]. If data security did not exist, it would almost be impossible to govern different aspects of society. The data stored by various organisations and providers have sensitive information and should be protected. The law contains principles that governments, organisations, and businesses should adhere to protect consumer data by ensuring the practices do not compromise the accuracy, safety, and security of the data and ensure privacy. Data security plays a crucial role in keeping a check on cybercrimes by providing fundamental consumer data such as banking information, addresses, and contact information to remain protected. Breaches of database systems have proven to be costly, and the affected parties often pursue compensation against the organisations responsible for the violations. Failure to adhere to existing secure practices exposes organisations to punishments for non-compliance with cybersecurity. Violation of cybersecurity laws and regulations often results in dire consequences making adherence to cyber-security policies crucial for the success of organisations [BB16].

2.1.2 Regulating Personal Data

The increased regulation on data protection and privacy is fuelled by cybersecurity threats and concerns regarding new technologies. Data breaches have raised concerns about how secure personal data is. Technological innovations such as the Internet of Things, the introduction of 3GPP 5G[2], IPv6 and Big Data applications call for regulation.

The summarised term IoT is not a technology per se but describes the strive of connecting devices and applications [KKZK12] to a network of sensors and actors producing a large amount of data. The International Telecommunication Union (ITU) also mentions besides

[2] https://www.3gpp.org

their formal definition of IoT as a “global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies” [(IT12] that “IoT can be perceived as a vision with technological and societal implications” [(IT12]. Most of the devices could potentially be used to gain access to personal information, hence reinforcing the existing regulations [Glo14]. Whereas IoT devices, as ubiquitous or sometimes even wearable devices, create convenience for the users, they also produce large amounts of data that can be used for profiling. Also, IoT devices create new or aggravate existing security issues due to the lack of software updates, the high number of devices, the limited processing power and storage and therefore the need for new lightweight security algorithms on these devices [ZCW$^+$14]. The 5th generation technology standard for broadband cellular networks (short 5G) is an enabling technology for mobile IoT applications. It is the next generation of wireless mobile networks that, besides speed improvements, add reduced latency and the possibility of peer-to-peer communication between mobile clients and machines [BHL$^+$14]. Due to the reduced latency and higher bandwidth, it has the potential to move data quickly and therefore make clients ‘thinner’ and store data in centralised systems or cache “massive amounts of data at the edge of the wireline network right before the wireless hop” [BHL$^+$14]. The intentions of wireless 5G low-energy IoT devices are ubiquitous, online sensors with low energy consumption that allow user interaction either direct by speech or indirect by collecting data and semi-automated actions with the cloud.

This collection bears apparent risks. Past security breaches and personal data leaks such as Uber [ZZPZ19] exhibit just how concerned the public should be in terms of data protection and privacy in centralised data storages [Byg14]. Uber stores information concerning drivers, vehicle details, and consumer data, such as addresses and financial information. The information is invaluable to criminals who are becoming more innovative in their attacks. In the cases of Israeli voters’ information leak [VFK20], Buchbinder’s customer information leak [Sal20] shows that information is also leaked accidentally without any malicious third party involved. McCandless and Evans [ME21] collected and visualised known data breaches with over 30 000 data records since 2004, showing a rising trend. These issues were more a misconfiguration than a highly sophisticated attack on systems dealing with large amounts of personal data. Therefore, the increased regulations are aimed at securing public information assets by regulating involved bodies to ensure the security of the consumers. Regulations create an environment where companies are obliged to protect the information people entrusted them with. Regulations backed by punishments create a general and specific deterrence. Meaning they, on the one hand, influence the whole public by illustrating a punishment and, on the other hand, directly influence the offender by punishing him or her [VMH06, p. 57]. Both aim to reduce the number of incidents by regulating the allowed and punishing violations. Technological advances produce more data, make the transfer easier, and the collection and interpretation of it more valuable. Therefore regulations try to restore the balance between people and companies by imposing duties for companies and granting rights to people.

2.1.3 Personal Data and Information

Personal data refers to those data or, consequently, information that can be linked or traced back to an individual or a narrow group of individuals. The GDPR and CCPA do not distinguish between data and information; therefore, the term ‘personal data’ is used in general. Personal data constitutes the identity of an individual as well as the individual possession, including intellectual property, provided it can be traced back to the individual. Privacy concerns over personal data are at risk in the era of digitisation and rapid technological advancements as they form the basis for fraud and identity theft that can constitute serious consequences under a false identity [Cha17]. Personal data such as the date of birth, sexual orientation, religion, addresses, or the Internet Protocol address of a device, as well as other metadata, are some of the explicit forms of personal data. Implicit personal data includes the online behaviour of individuals on social media that can be linked to an individual. Personal data is categorised into sensitive and non-sensitive personal data.

Sensitive data is data of special value to the person that might harm the person when known to others. It is usually defined as an exhaustive list in laws. These lists tend to be narrow and do, by definition, not allow an extension by those interpreting the law. Another method of defining the sensitivity is by defining the purpose of the use of the data (purpose-based approach) or the context it is used in (contextualised approach) [Won07]. Both tend to include data that could be used in an improper manner, but making the decision whether it is sensitive data is hard for the publisher in advance. Profiling data (see Section 2.3) can, for example, be used in several ways when combined with other data.

Society has grown in complexity in terms of social classes and the role of different individuals in the community. Individuals holding essential positions in society, such as organisational and political leaders, are often defined as people with public personalities who must remain specially protected from external attacks and data breaches.

Laws have different definitions of personal data. The United States of America generally use the term of Personally Identifiable Information (PII) whereas the European Union defines ‘personal data’. Personal data in US law is information that can be linked through referential and non-referential ways to a person. However, US law mainly focuses on the referential method, where descriptions or attributes of a person are used in determining personal data. Personal data, therefore, constitutes all information that together can help in the identification of a natural person [KL14]. When comparing personal data to PII, personal data is a broad term that includes all PII, but also other data relating to an individual that could lead to an identification. The US law, as well as the GDPR is clear on a natural person meaning that it is an individual human that is alive. Data related to a deceased individual is not considered personal data. Single-member companies that allow direct identification of the natural person are, according to US law, not natural persons, but in GDPR. The used term ‘all information’ is inclusive of objective and subjective personal information. Objective information is descriptive, including the sex and appearance of an individual, while subjective information includes employment evaluations. Inaccurate information that can help in identifying a specific individual is considered personal information. However, erroneous information that does not help in identification is not considered personal data [RR14]. Personal data should be able to distinguish an individual from other individuals hence making them identifiable. The identifiers commonly used to identify individuals online include IP addresses, cookies, and other identifiers [ADB14]. Identifiers refer to data points that help in the accurate identification of individuals, hence constituting personal data.

The GDPR defines personal data as "any information relating to an identified or identifiable natural person (data subject); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors[...]" [Cou16a, Art. 4]. This means that all information that can lead to the direct or indirect identification of a natural person is considered personal data. The information consists of a wide range of data presentation formats that can potentially contain personal data.

The California Consumer Privacy Act defines personal data ('personal information') in a way that only California residents that are 'consumer' of a 'business' are in the scope of the law and that the information "identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household" [CCP]. It also does not define sensitive data or 'special categories of personal data' as the GDPR does but defines some special rules for DNA and biometric data.

Reality shows that only a small amount of not directly personal related information is needed to identify a person. Research [Swe00] showed that 87% United States citizens could be identified only with a given postal (ZIP) code, gender and date of birth and about half of them (53%) are likely to be uniquely identified by only place of residence, gender, date of birth.

The identifiability of a person in personal data categorises the quality of personal data further. Simple classifications consider personal and anonymous data, whereas anonymous data cannot be related to a person. Fine-grained definitions [PTF16] classify data in

- direct- and indirect identification data,
- pseudonymous data,
- de-identified data,
- anonymous data,
- and aggregated anonymous data.

Direct identification data identifies a person without the need for further data, for example, a full name and address. Indirect identification data needs further data in order to be able to identify a person, for example, an IP address or a car's Vehicle Identification Number. Pseudonymous data add a classification that relates to data that has lost its direct relation to a person by introducing some form of intermediate data (e.g. a pseudonym) and storing the direct-personal data separately. These are considered as a form of data protection whereas usually still considered as personal data. The de-identification techniques are often prioritised by corporations dealing with the processing of big data [BB17]. De-identified data is usually defined by the inability to identify a natural person from the data with reasonable effort. The main risk with de-identified data lies in the existence of unique characteristics of a person and the existence of external data sources that can be used to extend the data to help further identify a person [Ema13].

Anonymous data refers to information that cannot be retracted to an individual by the processor or any other person. Data is considered anonymous if re-identification efforts prove futile. Anonymised data is not considered personal data as it is challenging to identify an individual from the information on anonymous data. Data anonymisation techniques include the addition of noise, substitution, aggregation and differential privacy. Noise addition techniques allow data to be manipulated by adding imprecision or inaccurate information to the authentic data. Substitution occurs when the original values are replaced with new parameters [Bor16]. The technique is often combined with noise in addition to making the data more anonymous and difficult to re-identify the individuals.

Aggregation allows personal data not to be singled out by combining several data from different individuals sharing personal data. The technique ensures anonymity by omitting specific unique characteristics but leaving the data viable for future use [BB16]. Aggregation utilised k-anonymity that renders re-identification possible such as data suppression and generalisation. Generalisation amounts to l-diversity to reinforce the k-anonymity that can be lifted by interference attacks where the attackers reverse the visible values and access the real values [Bor16]. The l-diversity technique ensures data security by ensuring all attributes have unique values. Differential privacy is a technique where a third-party requesting data is handed anonymous data that cannot be used in re-identification. Pseudonymous data encapsulates a concept of data protection where de-identification of data happens, but the data remains personal. Pseudonymisation allows for re-identification in indirect and remote terms, which makes the data remain personal. Although pseudonymous data does not allow or directly disclose the identity of the involved subjects, it can be used in identifying an individual through association with additional data [PTF16]. Pseudonymous data, therefore, remain subject to data protection guarantees. However, the rule of law affecting pseudonymous data is underdeveloped. The identification of pseudonymous information is assumed to have less significance to the individuals. Moreover, pseudonymous data can be processed if the anonymisation of data is not possible but should be conducted under advanced technical data processing standards. Pseudonymisation does not hide all the identifying information from the data but reduces the links of the dataset with the original data linked to an individual [Cha17]. Cryptographic hash function functions and tokenisation are some of the techniques applied to the pseudonymisation of data. Unkeyed hashes are used to map data to a fixed-size digest and are easily computable. A cryptographic hash function must have all properties of an unkeyed hash function and, additionally, pre-image resistance, second pre-image resistance, and collision resistance [MvOV18, p. 323]. Pre-image resistance means simply that searching for a message with the same hash as my hash must be difficult. Second, pre-image resistance means that searching for a message with the same hash as my message must be difficult. Collision resistance means searching for any two messages with the same hash must be difficult.

Cryptographical hashing does not automatically make secure pseudonymous data, especially not anonymous data. Re-identification of data might be possible by several techniques, such as rainbow tables on low entropy datasets or linking information [Eur19a].

Tokenisation allows components of data to be substituted with non-sensitive equivalents with no exploitable values but serving as an identifier. The identifiers act as references that can be traced back to the original data and are referred to as tokens. Since it has an indirect relation to the identity, it is still personal data.

Health Insurance Portability and Accountability Act (HIPAA) defines methods of de-identification for medical data. They distinguish the needed skills of the attacker. An attacker is defined as an expert when it can "appropriate knowledge of and experience with generally accepted statistical and scientific principles" [Cen96, § 164.514 (b)(1)]. In contrast, HIPPA defines the rules under the safe harbour for non-expert attacks as the need that "residual information could be used to individually identify a patient" [oHHS12].

The solution against expert attacks is a sophisticated statistical analysis of the data in its diversity, whereas the non-expert de-identification only applies the removal of a list of defined identifiers.

2.2 Past and Present of Data Protection

Data protection has been in existence for a long time and seeks to protect the human right to privacy. The thread of the right to privacy is traceable from the ancient Greeks up to the declaration of Human Rights after World War II, as well as the introduction of the European Union's GDPR [Car18]. Privacy has undergone definitions that revolve around the state where an individual is not disturbed or observed by other people [TPRM18]. The Bible discusses secrecy and privacy on multiple occasions. Sirach 42 defines what a man should be ashamed of, including "repeating what you hear, and of betraying secrets" [CBNP10, Sir 42:1]. Noah's son Ham violated his privacy by sharing with his brothers that their father was drunk and naked in his tent instead of helping him. This resulted in anger and punishment by Noah [CBNP10, Gen 9:21]. Aristotle also had notions of privacy in discussions involving the distinction between the public sphere (polis) and the private sphere (Oikos), respectively. He saw the need for a man to be part of the polis but that each individual is in his own realm in the sphere of Oikos [DeC18, p. 10]. The Universal Declaration of Human Rights adoption 1948 was a turning point as it included the right to privacy in Article 12. The advancements in technology and the steady growth in information technology forced the legal frameworks to evolve and protect the right to privacy. The first Austrian data protection law, the DSG, came into effect in 1978. It already included the fundamental right to data protection but separated the private and public sectors very strictly [KAJS16]. Austria also was one of the first European countries with authority for data protection. In 1980 the OECD issued data protection guidelines as a response to the increased use and power of computers in data processing [Glo14]. The council of Europe also adopted the right to privacy through the data protection convection treaty 108 in 1981. The European Data Protection Directive 95/46/EC was implemented in 1995 and has the language and principles that have shaped the legal landscape of data security [KAJS16]. It was implemented in Austria with the data protection law DSG 2000. The GDPR, which informs most contemporary practices of data protection, was approved in 2016 by the European Union parliament. The regulation enforced in 2018 marks the most significant development in data protection in history. The model seeks to empower independent providers to undertake compliance with their fiduciary responsibilities of cyber-security. Cyber-security laws have evolved to not only try to address exclusive privacy but also reinforce additional concepts of fundamental scopes of privacy. The development is a significant achievement that seeks to restore control due to the rise in computer processing power as well as the extensive data available. Cyber-security has led to the creation of a data economy that seeks to acquire the trust and confidence of consumers. Democracy ensures that citizens are afforded essential rights and freedoms, including the right to privacy and data protection [ADB14]. Although a lot has changed from the ancient Greeks, only a few contextual things have changed over the years. Privacy and data security laws aim to ensure human happiness and peace by allowing individuals and their property to be treated with autonomy, dignity, and well-being.

2.2.1 Data Protection in the United States

The legal framework in the United States takes a different approach to addressing the issues of privacy, data protection, and associated cybersecurity threats. The country has sector-specific federal data protection laws that focus on specific types of data. The country generally lacks general federal legislation. For instance, the Drivers Privacy Protection Act (DPPA) of 1994 seeks to govern the use of personal information that is collected by state departments of Motor Vehicles [Ker16]. The act regulates the disclosure and privacy of personal information such as the identification number of the driver, name, medical information, disability information, and address. The Children's Online Privacy Protection Act (COPPA), in contrast, protects children's information at the federal level. The collection of any information from a child under 13 years is prohibited unless due to privacy notices and verifiable parental consent is available when the information from the children is collected. The video privacy protection act also protects the public from unauthorised and wrongful disclosure of videotapes and similar materials. The Cable Communications Policy Act of 1984 has provisions that aim to protect the privacy of subscribers. The federal and most state governments have legislation that prohibits recording communications from recording communications without the consent of one or all parties involved as per the involved statutes [DDFH$^+$15]. State laws have also been enacted to restrict and oblige businesses that relate to the collection, use, disclosure, security, and retention of specified information, including biometric data, medical records, driver's license information, email addresses, social security numbers, and library records. States also have discrete regulations concerned with surveillance issues, including cellular location tracking and drone photography. Different states have enacted various legislations that govern data breaches relating to the personal information of the residents. Businesses typically must comply with state laws, more so when it pertains to the personal information of the residents. Personal information is subject to definition by the laws of the state. Some states in the United States are more active and have strict data security. States such as Massachusetts have strict data security laws that require all entities dealing with the personal information of the state residents to implement, maintain and adhere to [Rau14]. New York cybersecurity regulations, for instance, govern financial institutions by setting the minimum standards for different compliance certifications [Ker16]. Other interesting sector-specific laws include the Gramm–Leach–Bliley Act (GLBA), also known as the Financial Services Modernization Act, which seeks to regulate how banks, insurance companies, and other financial service providers use personal information. The Fair Credit Reporting Act (FCRA) that was amended by the Fair and Accurate Credit Transactions Act (FACTA) also restricts the use of some categories of personal information to determine an individual's eligibility for credit, employment, and insurance. The FACTA and GLBA are the main laws in the United States regulating the collection and sales of personal data [(GA06, p. 17]. The Health Insurance Portability and Accountability Act (HIPAA) also seeks to ensure personal information is kept private and confidential. It is the main legislation in the United States legislation defining requirements for handling personally identifiable information collected during medical care and is split into two regulations, namely the HIPAA Privacy Rule and the HIPAA Security Rule [GI17]. The Telephone Consumer Protection Act (TCPA) and other associated laws and the Family Educational Rights and Privacy Act (FERPA) are other statutes that reinforce privacy and data security in the United States. There are many authorities in the country concerned with the enforcement of these acts, including the Federal Communications Commission (FCC), the Securities and Exchange Commission (SEC), the Consumer Financial Protection Bureau (CFPB), the Department of Health and Human Services (HHS), and the Department of Commerce. Data security and privacy in the United States is a developed topic, more so since the country is a technological hub and is bound to set the pace for data protection. The United States, in contrast, does not have a legal basis for individual protection. Although there is protection for several special groups and cases and a minimum level of protection through the Fourth Amendment, there is no overall comprehensive legal protection against the authorities and institutions.

2.2.2 Data Protection in Developing Countries

Data protection in developing parts of the world varies but is mostly informed by the concerned governments and moral reasons surrounding the protection of personal data. The global community understands that unrestricted access to personal information can inflict significant harm on the subject in different ways. Data security and privacy legislation, therefore, seek to prevent harm. Data security laws also help in consumer protection in contracts that determine how companies will use personal data [Glo14]. The governments and other private and public entities are often in a better position to ensure that the partners live up to the terms of the contract. The regulations help in establishing fair conditions for the contracts on personal data use and reduce informational inequality in the market. Personal data can also be used in different contexts and potentially culminate in informational injustice and discrimination of individuals. For instance, the medical history of an individual can disadvantage a business partner hence the need to ensure contextual integrity by enforcing data security laws [SABH15]. The world has also evolved to protect human beings against encroachment on moral autonomy and human dignity. The laws also seek to protect people from infringement of autonomy and protection of human dignity. The global acknowledgement of ethical reasons behind the protection of personal data ideally shapes data security in almost all countries around the world. The basic morality for the use of personal information is the need for consent when using personal data belonging to the subject. Individuals should be allowed to have control over data that could have potentially adverse effects on their lives [Flo14]. The processing of personal information should be for a specified purpose, and the use should be limited. Governments typically require that individuals should be notified when their data is used and allowed to correct the possible inaccuracies. The current privacy regulations undertake privacy by design framework.

China, as one of the largest countries, not only by size but also by population and economy, has a privacy-educational problem since the vast majority of the Chinese do not know what the concept of privacy is [Wan11]. The country is currently 30 years behind any western standards of privacy [Wan11]. Wang argues that China needs a further move to democracy in order to be able to adopt privacy and data protection. The latest advance in data protection is the Personal Information Protection Law of the People's Republic of China (PIPL). It is the first data protection law in China. It defines personal data, categories of data, legal bases and requirements for processing by companies and state agencies. Compared to the GDPR the PIPL only applies to processing operations within China, regardless of the origin of the data subject, in contrast to the GDPR that applies to activities of establishments within the EU, no matter where it takes place.

2.2.3 Future of Data Protection

Technological advancements are increasingly designed with data security in mind by incorporating privacy requirements in the operations radically, therefore, reducing the likelihood of privacy violations to occur [VVdB17]. Debates about data security and privacy around the world revolved around new technology, more so information technology. However, the era of globalisation has helped to standardise security protocols such that citizens living in liberal and democratic states enjoy their freedom and rights to privacy and data protection.

Data protection is deeply rooted in the need for data security, which is the safeguarding of crucial data from cyberattacks, data breaches, and access by unauthorised parties.

Data breaches violate privacy concerns in data as they entail unauthorised access to data to read or copy the information. The need for cyber-security came as a result of the realisation that stolen data containing confidential information could prove costly for organisations and consumers [BR17]. Data security and protection from cyberattacks on an organisational level are needed to protect personal data. Applications that do not produce, collect, and store data about individuals or only data that is essential for their activities are another option. Currently, businesses running these applications gain more than they lose by collecting arbitrary data when needed with consent by the user. Ben-Shahar describes the concept of data pollution [BS19]. He considers the current approaches of legal frameworks as focused on data and privacy protection instead of focusing on the real issue: "how the information given by people affects others, and how it undermines and degrades public goods and interests" [BS19]. He suggests a legal framework focused on controlling the external effects of data sharing.

Data transfers from one legislation to the other such as from the European Union to the United States, also called transatlantic data transfers, seem to be complicated in nature. Even though the intentions of both legislations are clear, a compromise that is legal in both legislations is apparently hard to find. The first attempt to build a legal base for data transfers to the United States was the 'International Safe Harbor Privacy Principles'. They laid a legal ground for US companies and EU companies to share private data when complying with the defined principles, namely notice, choice, onward transfer, security, data integrity, access, and enforcement [WA16, p. 5]. This agreement was declared invalid in 2015 by the European Court of Justice (ECJ). The second attempt was the 'EU–US Privacy Shield', which eventually was declared invalid by the ECJ in 2020.

2.3 General Data Protection Regulation

The introduction of a General Data Protection Regulation aimed to replace the EU Data Protection Directive 95/46/EC [Dir95], which was adopted in 1995. The idea was to unify and consolidate the data protection laws. In 2016, the EU regulation 2016/679 (also known as General Data Protection Regulation) entered into force and has been enforced since May 2018.

One core feature of the new regulation is the right to be forgotten [AA13], which means that a natural person has the right that his or her data be anonymised or deleted after the identification is not needed anymore for the purpose of the processing activity. Another more general feature is privacy by design which means that systems should be designed in a way to minimise the amount of personal data processed [Sch10].

With the European Union regulation (EU) 2016/679 General Data Protection Regulation (GDPR), data protection has received a legal foundation concerning the rights of natural persons (data subjects) whose personal data is processed [Cou16a, Art. 12-23]. With it also comes obligations of controllers (who decide the purposes and means of the processing) and processors of personal data [Cou16a, Art. 24-43]. The GDPR provides a large number of measures to protect personal data, which companies process, against misuse. This includes, among other things, the ability of data subjects to prevent further processing of their data, not only by the controllers but also by third parties. Furthermore, data processing should be made more transparent for the subjects by giving them a right to receive information about which data is processed for which purposes by whom and may at any time submit an application for modification or deletion of the data. The GDPR is applicable to all companies in the EU that process personal data, as well as third-country companies, if they offer services to EU citizens [Cou16a, Art. 3]. Transnational companies are only obligated to process data of European citizens according to the GDPR. Rossi notes that "for transnational companies, it is easier to adopt European Union regulations at home as well because if their products and services comply with the world's highest standards, they can be marketed anywhere" [Ros14]. Compliance with the data protection measures is enforced by public authorities, who also have extensive rights of access to the processing activities of personal data. Failure to comply with the data protection measures will result in fines of up to 20 million euro or, in the case of a business, up to 4% of its total worldwide annual turnover. In the case of a group of undertakings, the annual turnover of the entire group, not that of the individual legal entity, is considered.

2.3.1 Roles defined by the GDPR

The GDPR defines roles during the processing of personal data. Figure 2.2 shows the interaction between a selection of roles during a usual processing of personal data.

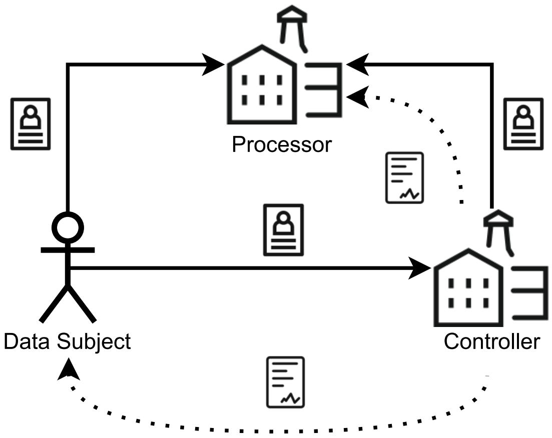


Figure 2.2: Relation of the roles in the European Union's General Data Protection Regulation

The 'data subject' discloses personal data regarding herself or himself in front of a 'controller'. The controller might transfer the personal data to a 'processor' or let the processor directly receive it from the data subject on behalf of the controller. The processor has a contract or 'data processing agreement' with the controller regarding the processing of personal data. The controller has ‘information obligations’ regarding the processing in front of the data subject.

The following sections explain the roles and their obligations and rights in detail. Recitals of the GDPR [Cou16a, Rec. 1-173], though not legally binding, are used for understanding the GDPR and applying it properly.

Data subject

A data subject is, according to the GDPR an “identified or identifiable natural person” [Cou16a, Art. 4 (1)]. A deceased natural person is not a data subject according to Recital 27 [Cou16a, Rec. 27]. An identifiable person is “one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” [Cou16a, Art. 4 (1)]. The indirect identification is important for the technical use of pseudonymisation and anonymisation. As discussed in Section 2.1.3 personal data can be categorised in terms of identifiability. Recital 26 [Cou16a, Rec. 26] clarifies that pseudonymisation is a technique used for data protection but does not have any influence on the classification of identifiability. Pseudonymised data is indirect identification data according to GDPR and does therefore require all the safeguards. The GDPR does not define de-identified data. It defines that anonymised data is data in which a data subject cannot be identified anymore, not by the controller nor by anybody else. The degree of anonymisation needed according to Recital 26 [Cou16a, Rec. 26] is defined by the possible effort (time and costs) needed in the future for re-identification. It is important to notice that neither tokenisation nor pseudonymisation (see Section 2.1.3) nor encryption is making data unidentifiable and, therefore, anonymous. Companies that are sole traders and, therefore, directly related to one natural person are considered data subjects according to the GDPR [LGO20]. Also, natural persons under 16 years are considered data subjects but are not allowed to consent to the processing of their information [Cou16a, Art. 8 (1)]. Again the GDPR defines that the controller must make reasonable efforts to verify that the processing is authorised or directly given by the guardian of the minor. The data subject has rights over various other roles in the course of data processing. Section 2.3.5 discusses the rights of a data subject.

Controller

The controller is a natural or legal person who “alone or jointly with others determines the purposes and means of processing personal data” [Cou16a, Art. 4 (7)].

The controller is the one who decides on the purposes and means, although there may be some freedom for a processor to determine the means based on the documented instructions from the controller. These include technical and organisational measures for data protection. The controller must have a contract or data processing agreement with the processor regarding the processing of personal data. The contract must contain the "subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller" [Cou16a, Art. 28 (3)]. The processor must also assist the controller with security audits and data protection impact assessment, amongst others, regarding the processing operations done by the processor and provide all information needed to demonstrate compliance [Cou16a, Art. 28 (3)].

The practical application of the interpretation of 'purposes' and 'means' has shown that the purpose criterion is more important than the means [Fin19]. The purpose of the processing must therefore be used primarily as a criterion for deciding whether a party is a controller. If a party determines the purpose of the processing, it is in any case qualified as controller [Par10]. Determining the means, in contrast "would imply control only when the determination concerns the essential elements of the means" [Par10]. The European Parliament Research Service (EPRS) called this notion 'primacy of the purposes criterion' [Fin19].

If both parties are defining the purposes and means of the processing operations, joint control can be assumed. [Cou16a, Art. 26 (1)] In this case, both are equally responsible for the processing operations and the resulting obligations. They shall define the according responsibilities regarding obligations and rights of the data subject to the aforementioned agreement or contract. The practical application has shown that joint-controllership has been used by courts, with the argument of ensuring the "effective and complete protection of data subjects" [Fin19]. In the case, Jehovan todistajat the CJEU ruled that "a natural or legal person who exerts influence over the processing of personal data, for his own purposes, and who participates, as a result, in the determination of the purposes and means of that processing, may be regarded as a controller within the meaning of Article 2(d) of Directive 95/46" [Cou18, par. 68].

Processor

The processor is a recipient of the data and is, in contrast to the controller, only partially responsible for the data processing obligations. The processor is usually a third party that provides data processing services, such as hosting, backup or printing. The controller must only work with contract processors who offer sufficient guarantees that appropriate technical and organisational measures are implemented so that the processing is carried out in compliance with the requirements of the General Data Protection Regulation and that the protection of the rights of the data subject is ensured [Cou16a, Art. 28 and Art. 30]. Even though the controller is responsible for selecting an appropriate processor, the processor does have legal obligations such as implementing security measures and assessing risks, but also supporting the controller in his obligations to the data subjects and in creating a data protection impact assessment and processing directory. Therefore, a contract must be concluded between the controller and the processor, specifying the subject and duration of the processing, the nature and purpose of the processing, the nature of the personal data, the categories of data subjects and the obligations and rights of the controller.

The contract shall also provide for the processor to assist the controller, as far as possible, with appropriate technical and organisational measures, depending on the nature of the processing, so that the controller can fulfil his obligations to respond to requests to exercise the rights of the data subjects [Cou16a, Art. 28 (3)]. The contract shall also define that the processor must assist the controller in complying with his obligations regarding data breach notification, obtaining a data protection impact assessment or prior consultation with the supervisory authority while taking into account the nature of the processing and the information available to her or him [Cou16a, Art. 28 (3)].

2.3.2 Legal basis for data processing

Every processing of personal data needs a legal basis according to [Cou16a, Art. 6 (1)]. When selecting the appropriate legal basis, it should be checked, as in the case of any data processing which is to be carried out on the basis of consent, whether there is a requirement for processing by the controller or its processors even if the consent is revoked by the data subject. The basis of consent should not be chosen in cases where the assessment comes to the conclusion that processing is necessary for the fulfilment of the contract or for the protection of the data subject's own legitimate interests or those of third parties.

However, the processing is permitted if there are legitimate interests of the controller or of a third party only if such processing does not override the interests or fundamental rights and freedoms of the data subject which require the protection of personal data [Cou16a, Art. 6 (1)].

The legal basis of consent can apply on a subsidiary legal basis as soon as there is no other legal basis on which the data processing can be based, in particular, the performance of a contract, legal obligation or legitimate interest of the controller or a third party.

Effective consent requires that the data subject has given his or her consent in connection with one or more specific purposes. The controller must demonstrate that the data subject has consented to the processing of his or her personal data. The consent also must be given prior to the start of data processing, although the consent does not have to be given in written form, but any affirmative action [Cou16a, Art. 4 (11)]. In the case of written consent given in the context of a declaration that also concerns other facts, the request for consent must be made in a comprehensible and easily accessible form in clear and simple language in such a way that it can be clearly distinguished from the other facts [Cou16a, Art. 7 (2)]. Consent must be given voluntarily, i.e. without coercion and without being linked in terms of making the performance of the contract dependent on consent which is not necessary for the performance of the contract, the so-called 'coupling prohibition' [Cou16a, Art. 7 (4)]. Obtaining consent does not relieve the controller of the obligation to inform when personal data is collected from the data subject (see Section 2.3.3).

If consent violates the requirements of the GDPR, the consent is invalid, which means that there is no legal basis for the processing, and it is inadmissible. However, [Cou16a, Art. 7 (2)] provides for the possibility of partial invalidity for consent if only separable parts of the consent are invalid; the non-data subject parts of the consent may, if they are separable in themselves and make sense, still be valid in isolated form.

2.3.3 Obligations of Controller and Processor

Controller and processor have obligations regarding their processing activities. Controllers, as the party deciding on the purposes and means of processing, have more responsibility and, therefore, more obligations regarding the processing operations. The controller is also responsible for only contracting processors that provide sufficient guarantees for appropriate technical and organisational measures. The GDPR still sets high standards on both controller and processor in terms of technical and organisational measures. These are needed to ensure the protection of the rights of the data subjects. The obligations of processors and controllers range from technical and organisational measures to information obligations towards data subjects and authorities. In special cases, it is also necessary to appoint a data protection officer and register with the data protection authority. The following chapters describe the obligations of processors and data controllers in detail.

Records and Transparency

The enforcement of transparency is one of the cornerstones of the GDPR. Information obligations towards data subjects are, therefore, very important. The data subject must, according to Art. 13 GDPR, be informed of the purposes, and the legal basis of the processing operations, the recipients and their rights regarding the processing operations, and others [Cou16a, Art. 13]. This information must be given to the subject wherever his or her data is recorded before the processing starts. The data subject must also be informed of the possibility of lodging a complaint with the data protection authority and of his/her rights (see Section 2.3.5). This information should help the data subject to make an informed and well-informed decision. Therefore, it is necessary that this information is provided in a form that is readable and understandable for non-technical users.

Nevertheless, controllers also have obligations to document and disclose information to the data protection authority if requested. According to Art. 30 GDPR, this includes information on the processing activities, including the purpose of the processing, the storage period of the data, the legal basis of the processing, the categories of data processed, and information on the transfer to third parties and third countries [Cou16a, Art. 30].

Table 2.1 summarises and simplifies the aforementioned recording obligations of the processor and the controller and compares them to the transparency rights of the data subject.

	Processor's records	Controller's records	Subject's information right
	Art. 30	**Art. 30**	**Art. 13**
contact details of controller/processor	2 (a)	1 (a)	1 (a)
data protection officer contact	-	1 (a)	1 (b)
categories of processing	2 (b)	-	-
purposes	-	1 (b)	1 (c)
legal basis	-	-	1 (c)
legitimate interests	-	-	1 (d)
categories of recipients	-	1 (d)	1 (e)
transfer to third countries	2 (c)	1 (e)	1 (f)
storage period	-	1 (f)	2 (a)
rights of the subject	-	-	2 (b-d)
contractual requirement	-	-	2 (e)
automated decision-making	-	-	2 (f)
technical and organisational measures	2 (d)	1 (g)	-

Table 2.1: Comparison of transparency required by the GDPR

Technical and Organisational Obligations

The GDPR requires technical and organisational measures to be taken by both controllers and processors. It allows for a weighing-off risk and costs but asks for the consideration of the current state-of-the-art and therefore updates. The aim of these technical and organisational measures is to secure the personal data being processed. [Cou16a, Art. 32] refers to techniques to be used, such as pseudonymisation and encryption (see Chapter 2.1.3), but also mentions the information security CIA Triad's properties (see Figure 2.3 and [And11]); namely, Confidentiality, Integrity, and Availability. The International Organization for Standardization (ISO) defines these terms in ISO 27000. Confidentiality, according to ISO is the "property that information is not made available or disclosed to unauthorized individuals, entities, or processes" [ISO18], and integrity as a "property of accuracy and completeness" [ISO18], and availability as a "property of being accessible and usable on demand by an authorized entity" [ISO18].

[Cou16a, Art. 32 (1) lit. b] requires the controller and processor to ensure ongoing confidentiality, meaning that data is only accessed by authorised parties in terms of reading data [And11]. It strongly relates to the concept of privacy, described in Section 2.1.1. The requirement for measures protecting confidentiality is needed in order to be able to trust a controller or processor to process data without spreading it unwillingly to third parties. In such a case, a data breach and an according plan must be executed (see

Section 2.3.3). Measures include protection of login credentials, physical and technical access and entry control, and logging of access and transfer. It is important to notice that these measures are not only technical but also organisational. Measures, such as access and entry control, must be implemented with organisational measures who has and had access and therefore implemented discharge management of a company and the general building access control management.

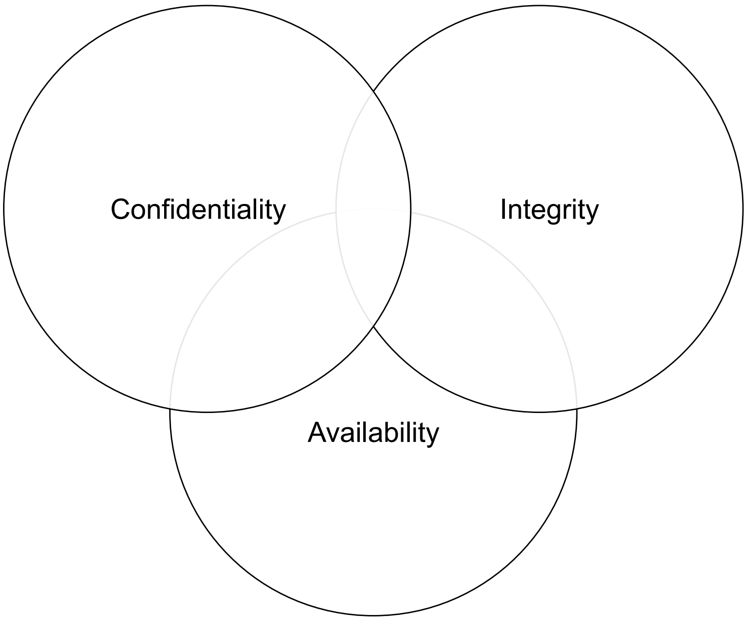


Figure 2.3: Information Security CIA triad (based on [And11])

[Cou16a, Art. 32 (1) lit. b] also requires the controller and processor to ensure the ongoing integrity, meaning that data is only changed by authorised parties in terms of adding, changing or deleting data [And11]. Transport security usually adds mechanisms for integrity, such as digital signatures that allow the receiver to check whether the data has been modified after being signed. Also, storage can be made fully or partially immutable, meaning that it can only be written once and not be changed (such as a data structure described in 3.3). Also, measures of confidentiality such as protection of login credentials, physical and technical access and entry control, and logging of access help protect data from violation of integrity. Also, mechanisms for reversing unintentional changes or deletions, such as backups or change logs, must be implemented. In case of an unauthorized or unintentional change of data, a data-breach notification must occur (see Section 2.3.3).

The deletion of personal data can also lead to the unavailability of data. [Cou16a, Art. 32 (1) lit. b] also requires the controller and processor to ensure ongoing availability, meaning that data is accessible when it is needed [And11]. As mentioned, availability can be the result of an integrity breach but also refers to any disruption of access, such as hardware or software failure and power outages. Also, Denial of Service (DoS) attacks can lead to a service being inaccessible as the result of personal data unavailability. Measures can be technical, such as DoS prevention, access control, and backups, but also organisational, such as power outage plans or disaster recovery plans.

[Cou16a, Art. 32 (1) lit. b] explicitly also mentions resilience in terms of being able to recover quickly from such incidents. Resilient systems can cope with failures of single components by designing so-called, High Availability (HA) systems. On an organisational level, resilience asks for continuous assessment and preparation for different scenarios. [BD05] describes a framework for resilient organisations and mentions activities and characteristics for organisations to be more resilient:

- **Risk management:** Reducing the likelihood that recoverable limits will be exceeded
- **Business continuity planning:** Moving the boundaries which define the recoverable limits for the organisation
- **Situational awareness:** Reducing the response time to recognise that change or action is needed
- **Creativity and responsiveness:** Improving the speed and capability of the organisation for responding to change

Plans help to quickly execute the prepared responses with pre-defined responsibilities and a budget for training and preparation. Recital 49 mentions Computer Emergency Response Teams (CERTs) and Computer Security Incident Response Teams (CSIRTs) related to the need to access data as a response to such incidents. These response teams gather experts that handle computer (security) incidents. These groups exist on the organisational, public and public-private levels. Austria has a National public-private partnership called 'GovCERT'. The Austrian NISG [Bun03] requires certain Austrian companies to report incidents [Bun03, § 19], implement technical and organisational measures [Bun03, § 17 Abs. 1], and to give the response team access to systems to verify compliance [Bun03, § 17 Abs. 4]. The chancellor is allowed to transfer personal data to the response team from any system affected by the law [Bun03, § 10 Abs. 6].

[Cou16a, Art. 25] defines measures to be taken upfront while implementing data processing processes and applications. It states the necessity to implement data protection by design and by default. Section 2.3.6 describes the implementation of these measures in detail.

Data Protection Impact Assessment

Organisations that are performing processing operations that are "likely to result in a high risk to the rights and freedoms of natural person" [Cou16a, Art. 35 (1)] need to do a data protection impact assessment. A data protection impact assessment shall be done before performing an evaluation of personal aspects of natural persons (profiling - see Section 2.3.6), which is subsequently used as a basis for decisions that could have legal effects on natural persons or affect them in a similarly significant way [Cou16a, Art. 35 (3) lit. a], or when processing a large scale of data from special categories (see Section

2.1.3) [Cou16a, Art. 35 (3) lit. b], or systematic monitoring of a publicly accessible area [Cou16a, Art. 35 (3) lit. c]. National implementations such as the Austrian data protection law ‘Datenschutzgesetz’ (DSG) define black and white lists that need to be considered. When assessing if a data protection impact assessment needs to be done.

The impact assessment must, according to [Cou16a, Art. 35 (7)] at least contain a description of the processing operations and the purposes of the processing, including the legitimate interest pursued by the controller, which is already required by [Cou16a, Art. 30] (see Section 2.3.3). Furthermore, it needs to contain the assessment of the necessity and proportionality of the processing operations in relation to the purposes [Cou16a, Art. 35 (7) lit. b], and an assessment of the risks to the rights and freedoms of data subjects [Cou16a, Art. 35 (7) lit. c]. These risks then need to be addressed by implementing safeguards and security measures.

The rights and freedoms are, according to the opinion of the Article 29 Data Protection Working Party, not limited to data protection and privacy but might also include “other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion”[Par17a].

The data protection impact assessment guarantees that risky processing operations are assessed thoroughly in order to lower the risks for the processing organisation and, ultimately, the data subjects concerned.

Data-breach Notification

Where personal data was accidentally or unlawfully destructed, lost, altered, unauthorised disclosed, or unauthorised access to, personal data unauthorised transmitted, stored or otherwise processed a data breach according to [Cou16a, Art. 4 (12)] occurred. In case of a personal data breach, the controller needs to notify the data protection authority and the data subject affected. [Cou16a, Art. 34 (3)] lists exceptions in which the controller does not have an obligation to notify data subjects: If the data is not readable by anyone because of technical or organisational protection, such as encryption; If the controller has taken measures to mitigate the “high risk to the rights and freedoms of natural persons”; If the effort is disproportionately high to individually contact data subjects, the controller can publicly publish the communication. The data protection authority checks whether the above criteria apply.

The data breach notification to the data protection authority must include all information shown in Table 2.2. Data subjects need to get information in clear and understandable language and do not need to have information to categories and the number of data subjects or categories and the number of records affected.

Appointment of a Data Protection Officer

Controllers and processors must appoint a data protection officer if they process personal data as a public authority or body (with the exception of courts) or if their main activity

	Communication to DPA	Communication to data subject
	Art. 33	Art. 34
categories and approximate number of data subjects	3 (a)	-
categories and approximate number of personal data records	3 (a)	-
the nature of the personal data breach	3 (a)	2
contact details of the data protection officer	3 (b)	2
consequences of the personal data breach	3 (c)	2
measures taken to mitigate adverse effects	3 (d)	2

Table 2.2: Comparison of communication in the case of a data breach required by the GDPR

consists of monitoring data subjects, or if their processing involves special categories or criminal convictions and offences on a large scale [Cou16a, Art. 37 (1)].

The data protection officer may be an employee or an external service provider. He or she should be involved at an early stage in the processes related to the processing of personal data. These processes include the introduction of new processes, the ordering of hardware or software, and the tendering of a contract. His or her responsibilities are defined in [Cou16a, Art. 39 (1)] as:

- to inform and advise the controller or the processor and the employees who carry out processing activities of their obligations
- to monitor compliance with this regulation
- to raise awareness, training of staff involved in processing operations, and the related audits
- to provide advice where requested as regards the data protection impact assessment and monitor its performance
- to cooperate with the supervisory authority

The obligation and existence of the data protection officer in companies are important in order to be able to ensure control and compliance with the data protection regulation.

2.3.4 Purposes of the data processing

The processing of personal data needs one or more defined legitimate purposes [Cou16a, Art. 5 (1)]. A simple indication that personal data will be collected and processed is not sufficient. This basic principle of the GDPR is known as the ‘purpose limitation’.

Personal data must only be stored while having a valid purpose. It must be stored in a form which permits the identification of data subjects only for the time necessary for the purposes for which they are processed on the principle of data minimisation (see Section 2.3.6).

A stated purpose must answer the question of why the data subject's personal data is processed in order for the data subject to be able to make an informed decision regarding the processing and the implications thereof. The period for which the personal data is necessary for the purposes of the processing must be defined and documented (see Section 2.3.3).

In cases where processors are involved, in particular [Cou16a, Art. 28] and [Cou16a, Art. 30 (2)] are relevant. In particular, the controller must only work with contract processors who offer sufficient guarantees that appropriate technical and organisational measures are implemented in such a way that the processing is carried out following the requirements of the GDPR and that the protection of the rights of the data subject is ensured.

Thus, a contract must be concluded between the controller and the processor, specifying the subject and duration of the processing, the nature and purpose of the processing, the nature of the personal data, the categories of data subjects, and the obligations and rights of the controller [Cou16a, Art. 28 (3)].

The contract shall also provide, among other things, for compliance with the data security measures referred to in [Cou16a, Art. 32], for the processor to assist the controller, as far as possible, with appropriate technical and organisational measures, depending on the nature of the processing, so that the controller can fulfil his obligations to respond to requests to exercise the rights of the data subject referred to in [Cou16a, Art. 12-23].

The contract must also include the obligation of the processor to assist the controller in complying with his obligations regarding data breach notification, obtaining a data protection impact assessment, depending on the data being processed, the processing activities and the information available to the processor. A processor must also keep a register of processing activities carried out on behalf of the controller.

2.3.5 Rights of a data subject

The principles of data protection stated by [Cou16a, Art. 5-11] long for fair and transparent processing. Under the GDPR, all data subjects have a number of rights with respect to their personal data. These are, amongst others, implemented in the rights of the data subjects [Cou16a, Art. 12-23]:

- right to information and access;
- right to rectification;
- right to object and restrict processing;
- right to erasure of personal data;
- right to data portability;
- rights regarding automated processing and decision making.

In addition, data subjects have a right to complain if they believe that their rights have been violated. The right to be informed is perhaps one of the most important rights afforded to data subjects under GDPR. This right entitles individuals to be fully aware of how their personal data will be processed and for what purpose. Data controllers must provide clear and concise information about data processing activities in a way that is easy for individuals to understand. This information must be provided in a timely manner and free of charge upon request from the individual. The right of access allows individuals to obtain copies of their personal data as well as information about how it is being processed. Data controllers are required to provide this information free of charge and within one month of receiving a request from the individual. The right of access also enables data subjects to request rectification or erasure of any inaccurate or incomplete personal data held by the controller and also allows them to object to processing activities carried out on their personal data. The right to protect personal data can be exercised where there are grounds for believing that an unauthorized person has accessed or otherwise used someone's personal data in a manner that poses a threat to the safety or well-being of the individual. This may involve requesting information from the controller about measures taken to safeguard data, including pseudonymization and encryption.

The following sections discuss the specifics and implications of these rights.

Right of objection

The right of objection [Cou16a, Art. 21] is a right that enables individuals to object to the processing of their personal data. It gives the data subject the opportunity to decide whether or not his or her data should be processed and ensures that he or she has control over how his or her data is used.

The right of objection is an important tool for protecting the privacy of individuals. It gives people the ability to exercise control over their personal data and prevents it from being used without their consent. By allowing people to object to the use of their data, [Cou16a, Art. 21] helps to ensure that individuals can trust companies with their information.

The right of objection also plays an important role in protecting freedom of expression and freedom of assembly online. It allows people who do not want their personal data processed by companies or organisations to withhold that information, which can help them stay anonymous when participating in online activities such as: posting comments on websites, joining online discussions, or signing petitions.

[Cou16a, Art. 21] provides for the right to object to the processing in different cases:

- objection to processing operations on the basis of public interest or relating to the legitimate interests of the controller or of a third party;
- opposition to processing for the purpose of direct marketing;
- opposition to processing for scientific or historical research purposes or for statistical purposes.

In the event of an objection pursuant to ‘public interest’ according to [Cou16a, Art. 21 (1)], further processing is only permissible if the controller proves compelling reasons for processing worthy of protection which outweigh the interests, rights and freedoms of the data subject or if the processing serves to assert, exercise or defend legal claims.

In the case of direct marketing, processing and related profiling must be stopped. In the case of ‘scientific or historical research purposes or for statistical purposes’ according to [Cou16a, Art. 21 (6)] the processing also has a counter-exception which requires the performance of a task carried out in the public interest.

By providing a justified objection, the data subject may prevent the further use and processing of the data concerned. However, a request for objection does not include the obligation to delete; this would have to be submitted separately as a request for erasure (see Section 2.3.5).

Right of data portability

The right to data portability according to [Cou16a, Art. 20] can also be exercised by the data subject during a valid contractual relationship. It covers data the data subject has ‘provided’ to the controller according to [Cou16a, Art. 20]. The data protection working party [Par17c] distinguishes data that has been

- actively and knowingly been provided,
- provided unknowingly by the use of a service or device.

They also clarify that information inferred from the data is not subject to the right of data portability.

The data subject has the right to receive these data in a structured, common and machine-readable format and has the right to transfer these data to another responsible party. This is subject to the condition that the processing is carried out either on the basis of consent or explicit authorisation concerning specific categories of personal data or the performance of a contract, in each case, by automated means. The right of portability is not provided in the case of processing for other reasons, nor for data which have not been collected from the data subject.

Correct classification of the purpose according to [Cou16a, Art. 6] for the specific data processing must be made. A system that is only intended to meet the minimum requirements of the right to data portability would have to ensure that the legal bases are known and correctly recorded and the origin of the data is known so that only the personal data provided for is made available.

The controller is also obligated to make personal data within the scope of [Cou16a, Art. 20], which originate from other sources or which are processed on the basis of other legal bases available to the data subject. No obligation of secrecy can be derived from [Cou16a, Art. 20 (3)], which implies a prohibition of the transfer of personal data from processing operations on any other justification.

The right to data transfer does not apply to personal data of third parties contained in the data records of the data subject; these would have to be removed before transfer.

The data portability also allows the request that data be ported from one party to another. In this case, the receiving data controller is responsible for ensuring that the data provided are relevant and not excessive for the new purpose of data processing on the receiving controller side [Par17c].

Right to restriction of processing

The fundamental right to restriction of processing under [Cou16a, Art. 18] is based on the principle that data subjects should have control over their personal data. Although this right has been enshrined in EU law for some time, the GDPR enhances and extends it. If a subject exercises its right to restriction of processing, the controller may only store the data in principle, and further processing is only allowed in exceptional cases. These cases, in particular, are an additional consent by the subject to further process the data; the need for the data to be processed for the establishment, exercise or defence of legal claims; the need for the protection of the rights of another natural or legal person; for reasons of important public interest of the country[Cou16a, Art. 18 (2)].

The restriction on a system asks for technical or organisational measures to restrict further processing. This includes, by definition of the term ‘processing’ [Cou16a, Art. 4 (2)], the deletion of the data. It may also be necessary, for example, to actively remove published personal data, which is in particular complicated in distributed ledgers.

The data subject has the right to request the controller to restrict the processing if one of the following conditions is met [Cou16a, Art. 18 (1)]:

- the accuracy of the personal data is contested by the data subject for a period enabling the controller to verify the accuracy of the personal data
- the processing is unlawful, and the data subject opposes the erasure of the personal data and requests the restriction of their use instead
- the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims
- the data subject has objected to processing pursuant to [Cou16a, Art. 21 (1)] pending the verification of whether the legitimate grounds of the controller override those of the data subject

With regard to the case of restriction in connection with the assertion, exercise or defence of legal claims, it is pointed out that these are legal claims that the data subject has against the controller. Otherwise, the data subject would have the possibility of forcing the controller to retain data that he or she needs for legal prosecution against third parties, which is not in line with the system of the GDPR.

Right to erasure

According to [Cou16a, Art. 17], the data subject has the right to obtain from the controller the erasure of personal data concerning him or her if one of the following reasons applies:

- Achievement of purpose: the personal data is no longer necessary for the purposes for which they were collected or otherwise processed.
- Revocation of consent: The data subject withdraws the consent on which the processing is based, and there is no other legal basis for the processing.
- Opposition: the data subject objects, and there may be no overriding legitimate reasons for processing (if relevant).
- Unlawful processing: personal data is processed unlawfully.
- Provided for by law: The deletion of personal data is necessary to comply with a legal obligation under Union law or the law of the Member States to which the controller is subject.
- Collection of data from children: Personal data has been collected in relation to information society services offered (i.e. distance selling of electronic services).

The right of deletion refers to the level of personal data and not to the level of data processing activities or data applications. Therefore, a system must also ensure that the verification of whether a request for erasure is justified can be based on personal data and that it is possible to systematically check which data processing operations are concerned and whether there is another legal basis for the processing.

Erasure must be irreversible and must prevent personal data from being accessed and information from being obtained. Erasure extends to all copies. This obligation also applies to backup copies without distinction. Moreover, subsequent anonymisation or pseudonymisation is not considered sufficient if the data can be attributed to a specific person again through the use of additional information. In the case of immutable backups, a mechanism to delete data when restored is sufficient according to the French DPA CNIL [ndledl20].

Erasure must be carried out without delay [Cou16a, Art. 17 (1)]; however, for systematic reasons, it must be assumed that the responsible party has a certain amount of time to check the identity of the data subject and the reason for deletion. The data subject must be notified in case the erasure takes more than one month [Cou16a, Art. 12 (3)]. The maximum time to be taken is defined as three months after receiving the request [Cou16a, Art. 12 (3)].

Right of access and rectification

While the right to information under Articles 13 and 14 starts with the collection of data, the right to information under [Cou16a, Art. 15] is about general rights regarding whether and which personal data of the data subject are processed. The controller is responsible for providing this information.

The right of access [Cou16a, Art. 15 (1)] is the data subject's right to obtain from the controller confirmation as to whether or not personal data concerning her or him are being processed and, where that is the case, access to the personal data and information specified in [Cou16a, Art. 15 (1)].

The right of access allows individuals to effectively exercise their rights under the GDPR. They can, for example, check whether their data is being processed correctly and, if necessary, request its rectification. They can also exercise their right to restrict processing or object to processing altogether in certain circumstances (see Section 2.3.5 and 2.3.5). Finally, they can obtain a copy of their personal data by making a request for access to their data.

If a request is mistakenly addressed to a processor, the processor is not obliged to act, but it is recommended that the processor forwards the request to the controller. This applies not only to the right of access but to all data subject rights in general. In addition, the processor is obliged to assist the controller.

The controller should use reasonable means to verify the identity of a data subject who requests information, especially in the context of online services. In cases where the controller processes a large amount of information about the data subject, the data subject may be asked to specify the information or processing operations to which his or her request for information relates.

Organisations must provide individuals with copies of their personal data free of charge and in a commonly used electronic form unless this would adversely affect the rights and freedoms of others, such as trade secrets. The possibility of charging for further copies exists only in the case of manifestly unfounded or excessive exercise of the right of access [Cou16a, Art. 12 (5)]. The organisation may charge reasonable fees that are proportionate to the administrative costs associated with providing this information.

The information, which in principle must be provided within one month [Cou16a, Art. 12 (3)], must, according to [Cou16a, Art. 13] contain:

- the personal data processed and their categories
- the processing purposes
- the recipients or categories of recipients
- if possible, the planned duration for which the personal data will be stored or if this is not possible, the criteria for determining this duration
- all available information on the origin of the data, if the personal data is not collected from the data subject
- Information on profiling and meaningful information on the logic involved and the scope and intended impact of such processing on the data subject.

Furthermore, the controller must inform the data subject of the rights of rectification, erasure, limitation or objection to such processing and of the existence of a right of appeal to a supervisory authority. The right to rectify inaccurate data relating to the data subject under [Cou16a, Art. 16] also includes the right to complete incomplete personal data.

2.3.6 Data security measures

Data protection is on an organisational level but also on a personal level depending on security. To protect individuals when entrusting their personal data to a third party, the GDPR asks for technical and organisational measures [Cou16a, Art. 32]. Data security provides sufficient technical measures or improvements to protect data adequately against loss or manipulation. This section lists and describes the relevant findings in the context of the GDPR and data security.

The GDPR describes the required security in the processing of data, i.e. data security. Taking into account the state of the art, the implementation costs and the nature, scope, circumstances and purposes of the processing, as well as the different probability of occurrence and severity of the risk, the article calls for the following measures [Cou16a, Art. 32]:

- the pseudonymization and encryption of personal data
- the ability to ensure the confidentiality, integrity, availability and resilience of the systems and services related to the processing on a permanent basis
- the ability to rapidly restore the availability of and access to personal data in the event of a physical or technical incident
- a procedure for regular review, assessment and evaluation of the effectiveness of the technical and organisational measures to ensure the security of processing

The state of the art refers to the processes, equipment or methods of operation available. This means that it is the best performance of measures available on the market to protect IT security objectives and data security. In continental Europe, the EU Directive 2016/1148 [Dir16] of the European Parliament and of the Council from 6st of July 2016 concerning measures for a high common level of security of network and information systems across the Union is a European reference that is implemented in local laws. It does not have any practical implementation guidelines. This was done with the EU Cybersecurity Act. It introduces an EU-wide cybersecurity certification framework for ICT and strengthens the European Union Agency for Cybersecurity (ENISA).

The US American NIST Cybersecurity Framework is a framework defined by the National Institute of Standards and Technology (NIST). It was first published in 2016 and extended to Version 1.1 in 2018. It implements a similar guideline for the United States.

A comprehensive and practical guideline has been created by the German Federal Office for Information Security (BSI). The IT baseline protection (German: IT-Grundschutz) is a good practical collection of methods, processes and procedures as well as approaches and measures for various aspects for IT security objectives and their measures that are customary on the market [RH20].

For the purposes of complying with the GDPR, it is recommended to establish and further develop a security management system within the framework of data security measures. This requires an information security policy that regulates security-related goals and strategies as well as internal responsibilities and measures. This includes, in particular, the identification and evaluation of existing security risks and the definition of organisational and technical security measures.

[Cou16a, Art. 25] essentially regulates two aspects, on the one hand, data protection through technology design (Privacy by Design, [Cou16a, Art. 25 (1)]), and on the other hand, data protection through data protection-friendly default settings (Privacy by Default, [Cou16a, Art. 25 (2)]).

Within the framework of the data protection-friendly presettings according to [Cou16a, Art. 25 (2)] (Privacy by Default), the controller must take appropriate technical measures to ensure that only personal data whose processing is necessary for the respective processing purpose is processed by presettings. This is needed in order to protect users who either do not have the time or technical abilities to understand the processing operations and implications thereof. These presettings include “the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility” [Boa19]. For applications, this is typically implemented within a configuration interface or an accordingly preconfigured opt-in interface, separating the processing activities and aforementioned presettings. Organisational processes and measures regarding data protection must implement privacy by default as well [Boa19].

Privacy by design means that systems should be designed in a way to minimise the amount of personal data processed [Sch10]. There is a thin line between privacy by default and design since optional minimisation is one of the goals of privacy by default. Minimisation in privacy by design describes not only the measures of minimisation in terms of collecting only needed data from a data subject but also measures such as:

- Minimising the processing of personal data
- Fastest possible pseudonymisation
- Transparency with regard to the processing of personal data
- Possibility of monitoring the data processing operations carried out regarding the data subject
- Encryption of personal data
- Separation of data collection and data processing according to the purpose
- Access minimisation (need-to-know principle)
- Organisational measure (for example, a deletion concept)

.

All these measures contribute to the goal of minimising the amount of personal data processed. As discussed in Section 2.1.3, encrypted data and, in general, pseudonymous data is a data measure but do not release the controller of technical or organisational measures.

Controllers must consider privacy by design during requirements evaluation of processes and systems. Controllers must choose systems and processors in their ability to comply with the GDPR. In the context of privacy by design, the measures for implementing the data protection principles of data minimisation are based on the state of the art, the implementation costs and the data protection risks for the persons concerned. As a result, it is necessary to carry out a risk assessment on the one hand and a cost-benefit analysis on the other hand. The measures in question must be implemented even before processing starts, as they include, in particular, the planning, creation and installation of the system with which the data processing is to be carried out. In this context, guarantees in the processing must also be included, which serve to implement the requirements of the GDPR and local laws and to protect the rights of the data subjects.

In the context of weighing up the proportions, four aspects, in particular, must be considered, which are only vaguely described in the GDPR:

- the state of the art;
- the costs of implementing the measures concerned;
- the scope, circumstances and purposes of the processing;
- the risk involved in the processing.

In practice, this means that before starting a processing operation, the system by which the data processing is to be carried out should be analysed, and, if necessary, changes should be implemented in advance.

Pseudonymisation can be a measure to limit the risk of data loss or loss of confidentiality; an analysis of the risks must also consider whether an attacker can also gain access to data with which he or she can remove the pseudonymisation.

The demand for transparency and the possibility of monitoring by the data subject is related to the necessary information and disclosure obligations towards the data subject and corresponding documentation by the responsible party. Furthermore, data should generally be encrypted as far as possible.

One measure would be a consistent separation of data on the basis of the purpose of collection and data processing; if this is not possible, data could already be marked with a mark indicating the purpose at the time of collection, which must be considered in any further processing of the data. Such measures should not take place only within the controller but should start earlier, in particular when the data is collected or first processed.

An important consequence of the data economy is that a deletion concept must be drawn up, including the resulting storage periods based on the specific purposes of use. The necessity of a deletion concept also arises from the right to deletion [Cou16a, Art. 17] or the principle to process only for specified, explicit and legitimate purposes [Cou16a, Art. 5 (1) lit. b].

The most crucial point is data minimisation, a principle that should be consistently applied by data processing systems. Also, in this context, it is necessary to assess whether the specific personal data is needed for the purpose to be achieved and for how long and whether the company's respective position needs this personal data.

The risk detailed in the processing can be reduced with data minimisation. The amount of personal data, therefore the risk, can be reduced by reduced processing (e.g. collection, transfer). The minimisation can be categorised as horizontal or qualitative and vertical or quantitative [Boa19]. Horizontal or qualitative data minimisation describes how detailed data is. By collecting, for example, only the needed data, the data property is impossible to be disclosed. Vertical or quantitative minimisation describes the number of entries being processed; by transferring, for example, only the needed data entries instead of a full dataset filtered later. Both categories can also be applied when data is split, to be stored separately for data protection reasons. Encryption separates the data in a way that the data can only be interpreted when knowing the key.

The technical-legal aspects of encryption as data protection measures are being discussed in the literature. Spindler [SS16] argues that encryption is a data protection measurement or a safeguard in terms of separation of data since the data alone cannot be read when encrypted but needs a key to be considered direct or indirect identification data (see definition in 2.1.3). Hence encrypted personal data would be considered pseudonymous data and, therefore, personal data according to Art. 4. This is supported by the mentioning of encryption in the context of the GDPR as a technical and organisational measure to secure data [Cou16a, Art. 32 (1)].

Martini and Weinzierl [MW17] argue that the deletion of a pseudonymous identifier makes the data personal sufficient, referring to the inability of direct or indirect identification stated in Recital 26.

Anderl and Schelling [And20] argue that encryption can be used for the fulfilment of the right of deletion (see Section 2.3.5) since the key can be deleted and encrypted data can therefore not be read anymore, which would be a 'factual deletion'. This would especially help for the fulfilment of rights of data subjects where the data is not only stored with the controller but with processors that do not have direct access to the key. A particular case is data made public, such as on public blockchains.

Piska and Bierbauer [PV19, p. 178] argue that a 'logical' deletion, in contrast to a 'physical' deletion, is not sufficient. They argue that personal data must be 'destroyed' according to a CJEU judgement [Cou17] or at least anonymised so that the data can no longer be retrieved using 'conventional procedures'.

Recital 26 [Cou16a, Rec. 26] clarifies the difference between pseudonymity and the degree of anonymity needed for anonymous data. It states that "all objective factors, such as the costs of and the amount of time required for identification" [Cou16a, Rec. 26] and the "available technology at the time of the processing and technological developments" [Cou16a, Rec. 26] need to be considered. Whether encryption will be secure in the foreseeable future depends on the key length and whether an attack is found. Both depend on stated 'technological developments' and must be considered. A particular case is data that has been made public after encryption; since the encryption cannot be changed once the data has been made public.

Profiling

Behavioural profiling is designed to help companies understand the characteristics of data subjects that would otherwise be unknown and, based on that, predict their behaviour as accurately as possible [Mic13]. Profiling and automated decision-making are driven by technological advances and data processing capabilities. They are used to create profiles of data subjects in order to make decisions based on that profile, such as advertising, price differentiation but also credit ratings for loans or anti-money laundering.

The GDPR defines profiling as “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular, to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;” [Cou16a, Art. 4 (4)]. The data subject gains rights such as the right to object [Cou16a, Art. 21 (1)] and information rights regarding the profiling [Cou16a, Art. 13 (2)].

[Cou16a, Art. 22] grants the right of the data subject not to be subjected to profiling or to an exclusively automated decision in certain cases. It states that “the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her” [Cou16a, Art. 22 (1)]. In principle, such automated decisions are not prohibited. However, in cases where this decision has a legal effect or where there is substantial adverse processing, this method of decision-making is prohibited. This provision is intended to prevent that, in principle, no such decision is made without a human decision step, in the sense that no natural person is involved that can choose between at least two options and can independently apply criteria. In order to avoid a case of an automated decision, a natural person would have to be able to freely make those decisions that have a legal effect on the person concerned or otherwise significantly affect him or her in a similar way. [Cou16a, Art. 22 (2)] lists cases where profiling and decision making is allowed, such as “necessary for entering into, or performance of, a contract between the data subject and a data controller” [Cou16a, Art. 22 (2)]; any processing in such way that it is authorised by Union or Member State law, including fraud and tax-evasion monitoring [Cou16a, Rec. 71]; and the consent of the data subject.

By ‘legal effect’, it is meant that a decision changes the legal position of the person concerned. This applies, in particular, to the decision of whether or not to conclude a contract. The EDPD also mentions the freedom to associate with others, vote in an election and take legal action [Boa18]. The term ‘similarly significantly’ is subject to discussion since the significance of an automated decision is debatable. The EDPD suggests that the significance of the effect must [Boa18]:

- significantly affects the circumstances, behaviour or choices of the individuals concerned;
- have a prolonged or permanent impact on the data subject; or
- at its most extreme leads to the exclusion or discrimination of individuals.

[Dro20] suggest that decisions that affect "dignity, integrity or reputation" [Dro20] are considered significant. In order to assess the significance of a decision-making process, the specific case must be considered with all its facets because the significance depends greatly on the circumstances and the data subject involved.

The GDPR does not contain a definition of an automated decision but defines what is meant by profiling. According to [Cou16a, Art. 4 (4)], profiling is understood to mean any automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular, to analyse or predict aspects relating to the performance of work, economic situation, health, personal preferences, interests, reliability, conduct, whereabouts or movements of that natural person. Examples of profiling in the literature include the following: automated procedures relating to shopping tips or purchase forecasts, the assessment of credit and solvency risks, the creation of personality profiles, geomarketing, or relating to driving behaviour.

Counter-exceptions exist if it is necessary for the conclusion or performance of a contract between the data subject and the responsible person, if it is provided for by legal provisions or if it is carried out with the explicit consent of the data subject.

A data subject that is affected by automated decision-making has the right to be informed of the automated decision-making or profiling [Cou16a, Art. 13 (2)]. It also has the right to get "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing" [Cou16a, Art. 13 (2)] if the decisions are based on special categories or produce "legal effects concerning him or her or similarly significantly affects" [Cou16a, Art. 22 (1)].

Furthermore, in cases of performance of the contract and consent, the data subject has the right to have a natural person appointed by the controller to review and, if necessary, amend this automated decision.

2.4 Discussion

It has been shown that data protection has many facets in various scientific fields. The interrelationships and tensions between these have been highlighted. One of the most deeply anchored aspects of this work is the tension between technology and legislation. Accordingly, data protection in the sense of data security but also technical and organisational measures were examined. In particular, the implementation in the European area by means of the GDPR was dealt with. The legislation provides the data subject with the means to rebalance the discrepancy between companies that process their personal data and the data subject. These means are, on the one hand, obligations of the companies and, on the other hand, rights of the data subjects. The obligations force companies to carry out an internal audit and to analyse processing activities and their appropriateness. Combined with the need to disclose these, this can already lead to a change. For the data subjects, it remains questionable whether the change in processing actually happens. By forcing companies to state the purposes of processing operations, transparency is at least guaranteed. In the case of WhatsApp's[3] extension of processing activities and transfer to their parent company, Facebook showed the results of transparency in this regard [Ovi21]. Due to the required transparency laws in the European Union, a public discussion about the changes took place. The division of the data protection conditions into a European version, and an American version showed the effects of the GDPR figuratively [Beh21]. The required technical and organisational measures, in contrast, help citizens to feel more secure when personal data is passed on to companies.

Technology itself is creating issues with advances that allow for more data to be analysed quicker and more information to be gathered from less data. Technology the same time, enables new solutions for privacy issues. The use of technology to support privacy in the sense of data-protected applications is a broad area. Privacy Enhancing Technologies (PETs) attempt to improve this field. These technologies are focused on finding implementations in the organisational environment that facilitate privacy-compliant implementations. PETs help to implement high data protection standards.

The following chapters will draw an image of one specific technology, namely blockchain. Like other technologies, it can contribute to data protection, but it also has its own problems with data protection. In order to be able to shed light on this in the following chapters, the relevant parts must be defined and understood in order to be able to draw scientific conclusions.

[3] https://whatsapp.com

CHAPTER 3

Distributed Ledgers and Blockchain

To discuss the data protection aspects of distributed ledgers and blockchains, in particular, the technology itself must first be understood. The following chapter, therefore, defines distributed ledgers and blockchain technology. The latter is described in detail with regard to its data structures and protocols. The chapter concludes with a brief summary of the current applications of blockchain technology in various fields.

3.1 Distributed Ledger Technologies

Distributed Ledger Technologies (DLTs) have seen an incline of public interest in recent years [Goo21]. The interest was fueled by speculation and the promise of getting rich quickly by investing early in the next fast-growing cryptocurrency. Blockchain is one of the most well-known distributed ledger technologies, but it is still vital to understand distributed ledger technology in general before examining blockchain technology in detail. Scientists and engineers have the duty to question the appropriate use of distributed ledger technology and make recommendations based on scientific methods. Reports and white papers can be interpreted as indicators, but especially in the case of distributed ledger technologies, the methods and interests of the authors must be strongly questioned. The media attention led to high investments in the field of distributed ledger technologies, which in turn led to an increase in the number of projects. While these projects usually had a well-articulated vision, they rarely had any real value; many of them were even scams [Gro21]. Prominent names became figureheads for big-postulated projects, often without the expected success. Initial Coin Offerings (ICOs), as a counterpart to Initial Public Offerings (IPOs) on financial markets, or hybrid Security Token Offerings (STOs), have been used as financing options for such projects. In order to present these ideas, which are usually not very well thought out, in a technically advanced way, the creation of a white paper became the usual course for ICOs. Some of these white papers are also published as peer-reviewed articles in journals, but most of them would not stand up to peer review, having a low information content and not adding any information compared to industry standards and other ICOs [FS19]. It is therefore essential to separate media reports, white papers and scientific work. This is difficult in the DLT environment, as the technology is being developed with a little scientific background in many areas, and the publications are somewhat lagging behind. Despite this, it is important to adopt an evidence-based, scientific approach to the topic.

The following sections present distributed ledger as a general technology and its characteristics, followed by a detailed explanation of blockchain and its technology stack. A discussion of methods to decide the appropriate use of blockchain technology completes the chapter.

3.1.1 Properties and Definition of Distributed Ledgers

A distributed ledger is an “append-only store of transactions which is distributed across many machines” [XWS19]. In the original meaning, a ledger is a central append-only collection of entries, such as sales transactions, expenses, or financial transactions and is stored in one location. The entries in these ledgers were taken in order to have one collection of correct entries. There is no need for an agreement on what entries are correct since there is only one copy of it. A distributed system has, according to Steen and Tanenbaum [TvS18] two characteristics: it is a collection of autonomous computing elements (nodes) and appears as a single coherent system to users. The nodes in these distributed systems are not necessarily owned by one party, and they do not trust each other but still want to provide a trustworthy service to each other [CV17]. Distributed ledgers use nodes to register, exchange, and synchronise transactions in their particular electronic ledgers, rather than holding the data centrally as in a traditional ledger [KTKT19]. To store consistent data on such a system, the data storage needs to be coordinated so that the state of the data can be altered and accessed.

One characteristic of a distributed ledger is that it has no central authority that decides over new data entries and, therefore, the need for a consensus on the entries and a quorum between the nodes [MGM^{+}17].

Another characteristic is the property to have ‘continuity’, meaning that entries are added to alter the state instead of modifying the existing data. A blockchain, in terms of data structure, is a specialisation of a distributed ledger that is also continuously growing, but in an append-only, chronological manner, summarised in data blocks (more details follow in Section 3.3). The use of a ‘chain of blocks’ to achieve a secure and valid distributed data structure is no general feature of distributed ledgers.

Other distributed ledgers use data structures such as Directed Acyclic Graph (DAG) either with blocks or single transactions [KS19]. Tangle, as used by IOTA[1] and Hedera

[1] https://iota.org

Hashgraph[2], are the most famous implementations in this field. Others [IP18] argue that blockchain sidechains use a specific data structure called a list of linked lists, sharing some properties of the blockchains based upon.

Walport [W+16] adds a weak characteristic that distributed ledgers are "typically public" and summarises the aforementioned characteristics with the following definition of distributed ledgers:

- a continuous ledger database;
- nodes need to reach a quorum for new entries;
- spread across multiple sites, countries or institutions;
- and is typically public.

Regarding the abstract system achieved by distributed ledgers, it is often argued that distributed ledgers are not only distributed but also decentralised (cf. systems theory). The following sections discuss the decentralised authority and distributed data properties.

3.1.2 Decentralised Authority

Since there is no central authority in distributed ledgers, consensus algorithms are needed in order to achieve a self-organising system in the form of a decentralised system. In a decentralised system, decisions are made by the system's nodes without centralised control, or processing [Joh99]. In a distributed system, nodes want to achieve goals together and realise that by coordinating by exchanging messages with each other [TvS18]. Distributed ledgers differ in terms of the used consensus algorithm [LLH+20]. They are generally used to find a consensus regarding a decision whilst keeping in mind that a participant might want to exploit the system. In the case of distributed ledgers, an important decision is whether a data structure, such as a block or a single entry, should be added to the distributed ledgers. Depending on the algorithm, other decisions need to be made as well. Section 3.4 explains different consensus algorithms in detail. To achieve decentralised authority, nodes of a distributed system need to collaborate and therefore need to be either directly or indirectly connected to each other. Indirect connections are achieved by routing connections through nodes. Thus, a node is both a receiver and a forwarder (router) of messages. Distributed ledgers use peer-to-peer connections between nodes to achieve connectivity. The degree of connectivity or connectedness between the nodes is a crucial measurement in such systems [Joh99]. The propagation of information in a dynamic system, where nodes can be added during operation and, therefore, without static routing, is hard. Distributed ledgers use overlay networks on top of other networks to span a network between nodes (see Section 3.4.1 for details). Distributed ledgers often use gossip protocols to propagate information to nodes fastly. Gossip protocols work virally. One node receives information and spreads it to multiple connected nodes. This allows for exponential information dissemination [IP18]. Gossip protocols create overhead in a network and are, therefore, only selectively used for important messages.

[2] https://hedera.com

Other distributed ledgers use distributed hash tables, such as Kademlia, to perform fast peer lookups and locate files or resources [LLH+20]. Through shared identifiers, nodes can calculate the distance, in terms of latency, between each other, and information can be shared efficiently since only a few nodes need to be contacted during the search for information [MM02].

3.1.3 Distributed Data

The aforementioned properties of distributed ledgers define a narrow border to several other systems and technologies.

Distributed databases share properties with distributed ledgers in terms of distributed data. Some also offer continuity with write-ahead logs, usually used for crash and transaction recovery. The difference is the lack of decentralisation. A distributed database is “distributed and replicated across a number of servers that together form a cluster” [TvS18]. They are usually either full or read-only replications. Full replicas allow writing and reading, whereas read-only replicas only allow for reading data. The nodes in such a system are usually not peers, meaning that each node has its predefined role and connections. The organisation and, therefore, the control of such a distributed database is highly centralised. Such distributed databases achieve high transaction rates on the one hand but are very static in structure on the other. Distributed Online Transactional Processing (OLTP) database clusters reach transactions of 600 000 payment transactions per second with a few server instances [PTAA16]. Also, distributed database systems usually allow for data security. Database security can be classified into layers [FERES+14]: authentication, authorisation, integrity, and auditing. They have, depending on the implementation, table, row, or field-based access rights that restrict reading, writing, deleting and updating separately. Some distributed ledgers do have additional authentication and authorisation capabilities. Most distributed ledgers are publicly accessible and can therefore be read by anyone. Auditing is combined with the continuity of the data layer. Integrity is usually an integral part of making changes to past blocks by malicious actors nearly impossible. Distributed Ledgers, including blockchain, have tried to overcome these shortcomings with different solutions and implementations (for example [Yak18]). A correct comparison between distributed databases and distributed ledgers is, therefore, impossible.

3.2 Blockchain

Blockchain and its surrounding technologies gained much attention recently. The term blockchain became known in the cryptographic context from Nakamoto’s white paper [Nak08] on a peer-to-peer electronic cash system and its implementation of Bitcoin. The popularity of blockchain has then been dependent on the popularity of Bitcoin and its payment system. The possibility of exchanging currencies, in terms of legal tender, into cryptocurrencies leads to a market which has been subject to volatility. Due to fluctuations in this exchange rate, Bitcoin became a popular speculative asset and thus received media attention and public awareness. Further developed blockchain technology, including advanced functionalities such as smart contracts, led to a broader field of application. Smart contracts are programs that are stored on the blockchain and executed in transactions [XWS19]. These smart contracts can create complex systems such as virtual organisations (Decentralized Autonomous Organizations (DAOs)), options markets or digital assets (tokens). As a result, ideas about where blockchain technology can be applied became very popular. Thus, the word blockchain led to a veritable hype. Quickly, there was a hyperinflationary use of the word blockchain, for example, to increase brand value, to obtain project funding, but also to use it to commit a crime. The ability to use smart contracts to create tokens led to an explosion in the number of different tokens. Because of the open availability of the source code of most cryptocurrencies, copies or copies with slight modifications of known cryptocurrencies also occurred. This, in turn, led to a gold rush, which is characterised by investing as early as possible in an unknown cryptocurrency, where one expects a high increase in the shortest time.

Differentiating between hype, science, and reasonable technical use is, therefore, very important. The following sections walk through the technologies surrounding blockchain, how blockchain is applied, and how to make a decision about whether to use blockchain technologies.

3.2.1 Properties and Definition of Blockchains

Blockchain is an implementation of a distributed ledger. Therefore, it inherits all the distributed ledgers' properties (see Section 3.1), implementing them in a specific way.

Definition 3.2.1 (Blockchain [XWS19])
A blockchain is a distributed ledger that is structured into a linked list of blocks. Each block contains an ordered set of transactions. Typical solutions use cryptographic hashes to secure the link from a block to its predecessor.

To be precise, a blockchain is only in a simplified form, a growing backwards linked list of blocks that hold data. These blocks are practically immutably linked using cryptography. Each block contains a cryptographic hash of the previous block header, metadata and a payload. But there are cases where multiple blocks can refer to the same or such as the genesis block to no block at all as their previous block, resulting in a non singly backwards linked list, which is automatically resolved (see Section 3.3).

These blocks are created by consensus nodes (see Figure 3.1) and spread over a decentralised peer-to-peer network. Consensus nodes validate the transactions and create blocks of multiple transactions. Each node in the network performs verification and validation of blocks, stores and relays blocks and transactions. In reality, nodes might perform multiple roles and store only partial or meta information. A node might, for example, store the full blockchain (full node) and act as a miner (consensus node) at the same time. Nodes that want to receive payments create addresses by calculating a private-public key pair and calculating an address from the public key. Consensus nodes usually store parts of the blockchain needed for the creation of a new block. The specifics of consensus nodes depend on the consensus mechanism used by the blockchain implementation.

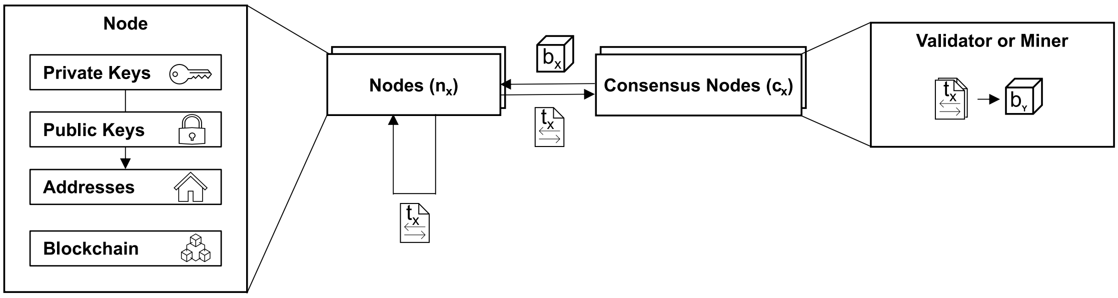


Figure 3.1: Generalised Blockchain Network Participants

Two well-known consensus implementations are Proof of Work (PoW) and Proof of Stake (PoS), where both have multiple implementations and sub-types. In blockchains using PoW, miners use cryptographic software and their computers' processing power to validate transactions and generate new blocks on the blockchain. They usually use specialised hardware for faster and energy-efficient calculations and form groups of miners where the profit is shared (mining pools) even though the computation could be executed on commodity hardware, but with little or no profit.

In PoS chains, it is the ownership that entitles a validator to be chosen to validate a transaction and collect a transaction fee for it. In the PoS concept, everyone who owns shares of the digital asset associated with the blockchain can become a validator and, depending on the shares, can theoretically validate up to a certain number of blocks without having to solve a computational problem on all nodes, in contrast to Proof of Work (PoW).

Blockchains can also be classified regarding their access restrictions. Public blockchains are open for everyone to read and have no read access restrictions; nodes can join and leave at will without needing permission [WWC$^+$18]. Writing to a public blockchain is, in contrast, restricted to consensus nodes (see Figure 3.1). This means that anyone with an internet connection and sufficient hardware can read and send transactions. It also means that anyone can become a consensus node for validation or mining of blocks. All nodes on the network can verify each new record added to the data structure (see Section 3.3), and the protocol includes a consensus mechanism to ensure the proper operation of the blockchain system, including the processing and inclusion of valid transactions in the ledger and the rejection of invalid transactions [XWS19]. This is usually safeguarded with economic incentives (see Section 3.4). The participants of public blockchains usually appear as pseudonyms, where participants can have multiple pseudonyms.

A private blockchain is restricted in a way that participation is only allowed for a few. The validation process is also restricted [WWC$^+$18]. Permissioned blockchains only allow certain actions for specific participants. Both require authentication and authorisation of the participants to allow restricted actions. This type of blockchain is usually run by an organisation with nodes optionally spreading over a consortium (sometimes called consortium blockchain).

In summary, blockchain stores transactions piratically irreversibly and, instead of a central authority, relies on a decentralised consensus that can be implemented in several ways.

3.2.2 Layered Architecture of Blockchain Systems

Blockchain Systems are systems consisting of nodes using blockchain as a data structure (see Section 3.3) with defined protocols (see Section 3.4).

Definition 3.2.2 (Blockchain System [XWS19])
A blockchain system consists of:

1. *a blockchain network of machines, also called nodes;*
2. *a blockchain data structure for the ledger that is replicated across the blockchain network. Nodes that hold a full replica of this ledger are referred to as full nodes;*
3. *a network protocol that defines rights, responsibilities, and means of communication, verification, validation, and consensus across the nodes in the network. This includes ensuring authorisation and authentication of new transactions, mechanisms for appending new blocks, incentive mechanisms (if needed), and similar aspects.*

The technology stack of a blockchain system can be conceptualised in layered models. These layers are independent and transparent to each other. A layered generalisation used in different blockchain system implementations helps to analyse, categorise and distinguish. Different [LWH20, LLH^{+}20, HVR^{+}20] architectural views exist, with different emphasis on functions of blockchain systems. Figure 3.2 compares three different layer architectures proposed in the literature.

The application layer is usually capable of definition and extension by more or less advanced programming languages. These programming languages stretch from non-Turing complete stack-based scripting languages to full-featured object-oriented programming languages. The application layer can be split into a virtual machine or interpreter that executes the application code. A contract layer [LWH20] distinguishes between the different layers of an application, such as the business logic or the interface defined in a contract layer and the user interface or application layer. Usually, a blockchain state is kept locally and is operated on by the interpreter or virtual machine executing the code. Homoliak et al. [HVR^{+}20] refer to the concept that a blockchain is a transaction-based state machine, especially in payment channels. To emphasise the importance of the state machine, they add a ‘state machine layer’ to their architecture. This is due to the fact that they have security considerations as an aim of their architecture.

The consensus layer of all shown layered architectures refers to mechanisms to find a distributed consensus, such as Proof of Work (PoW) and Proof of Stake (PoS) to decide on what data is included in a new block.

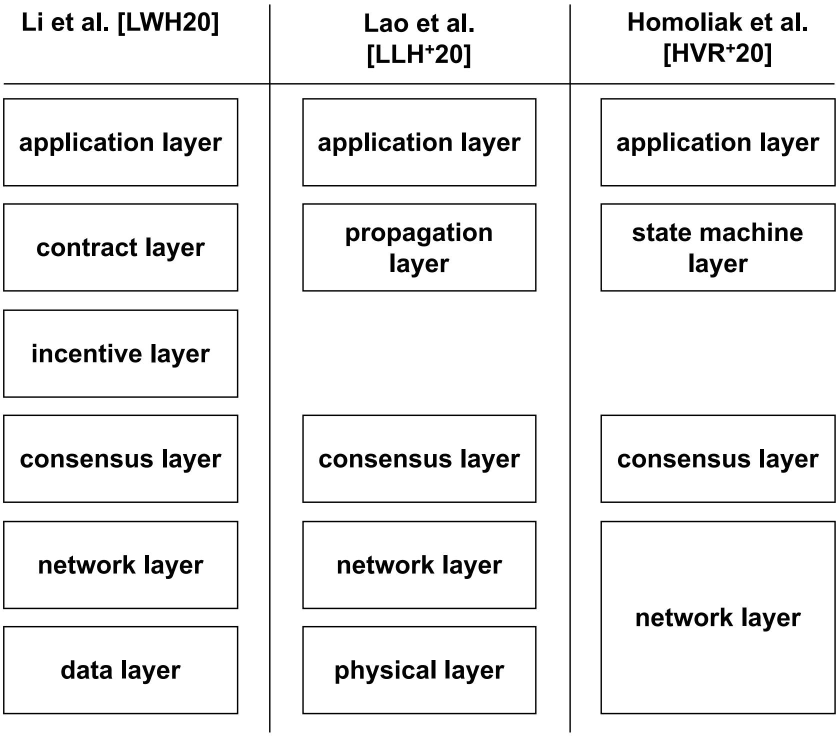


Figure 3.2: Proposed Layer Architectures in the Literature

The network layer implements the connection and discovery of nodes. They usually use a peer-to-peer overlay network (see Section 3.4.1) in order to exchange information in a decentralised way. Blockchain implementations usually have their own peer-to-peer protocol but use well-known mechanisms such as gossiping.

Li et al. [LWH20] add a non-technical layer called the ‘incentive layer’ to their architecture. This is due to the fact that the blockchain itself tries to steer the behaviour of users by incentives. A node should, for example, be incentivised to properly verify transactions and propagate them in the network instead of trying to selfishly calculate alternatives to the existing chains. The consensus algorithm can be seen as a game (cf. game theory). Blockchains assume that the implemented incentives are strong enough for a node to be fair instead of manipulating the game to its advantage.

These incentives are usually implemented by losing or gaining value, so-called release or allocation mechanisms [LWH20]. A blockchain, therefore, relies on a majority of fair players. The needed majority to withstand different attacks depends on the specific design.

The data layer is responsible for creating a data structure that fulfils the properties needed by the layers above. These properties are usually that the existing blockchain cannot be manipulated, data can be added to the data structure consecutively, and multiple single data entries such as transactions can be stored and found efficiently.

Lao et al. [LLH+20] add a physical layer to the architecture to emphasise the storage of the data structures. The data needs to be fastly accessible to verify and propagate transactions to other nodes. This is usually done with a database or other indexing mechanisms to find data in the physical layer. This layer also deals with the need for specialised hardware for implementing some proof of work implementations.

Some specialised blockchain implementations also try to fulfil networking capabilities such as low power for the Internet of Things (IoT) devices [FCFL18]. They, therefore, add physical network devices to the physical layer.

All layers build on each other and add distinct functionality to the blockchain. They are used to decide on which data entries and blocks are accepted on the chain, secure the integrity of blocks, propagate and store blocks, and interpret and compute based on the data with custom programs.

3.3 Blockchain Data Structures

A blockchain is, besides other considerations, a data structure. It is, in a very simplified version, a singly backwards linked list (more details later). Each list item is a block that again holds structured data. The backwards list structure is achieved by adding the identification of the previous to a block. The identification is usually a hash of the block (see Figure 3.3). This also guarantees the integrity of the whole chain. In certain scenarios, a block can have multiple ancestors or no predecessor. These cases are discussed in Section 3.4.2.

3.3.1 Blocks

A block is part of the blockchain and contains a limited number of transactions depending on its agreed size. The first block in a chain is called a ‘genesis’ block. Each block contains three pieces of information:

- a header with metadata, including a reference in the form of a hash to the previous block header;
- the number of transactions in the block;
- and a list of included transactions.

Figure 3.3 shows the usual construct of a blockchain on the example of Bitcoin. It is based on the documentation [Bit20] of Bitcoin core. The chain is a linked list, including a hash of the previous block header to make it immutable. The header of a block, therefore, includes a hash of the previous block header.

The nonce (and possibly other factors) are used to modify the block to change the hash outcomes of the mining process (see Section 3.4). ‘Nonce’ is an acronym for ‘number only used once’ and is a number that is found by miners when solving computational tasks (hash computation) and is part of the overall solution [AKR+13].

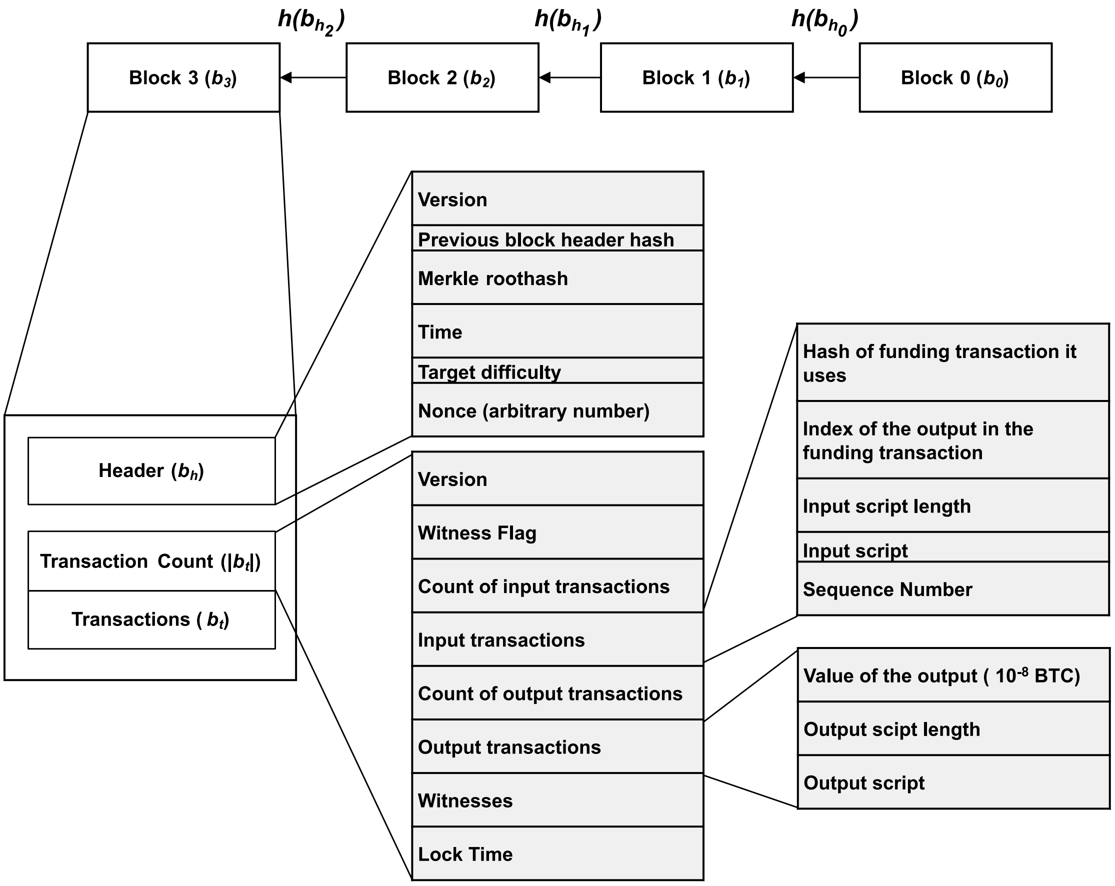


Figure 3.3: Structure of Blocks on the Bitcoin Blockchain

3.3.2 Transactions

The transactions themselves are stored in the transactions part of the block (see Figure 3.3). A single transaction can be funded by multiple input transactions. These input transactions prove the knowledge about the secret (private key) needed for using the outputs of previous transactions. The output transactions then define the spending of the transaction. Spending is defined by a simple scripting language. Each transaction itself, as depicted in Figure 3.3, contains an input and output script, depending on the type. These scripts are, in the case of Bitcoin, simple stack-based scripts that are themselves not Turing complete because of the absence of loops or other equivalent constructs, which are needed to guarantee termination.

This scripting language allows several standardised types of transactions. ‘Pay to public key’ transactions, for example, use the recipient’s public key to transfer funds. Whoever can prove to know the matching private key by signature can access this transaction’s funding but are deprecated. ‘Pay to public key hash’ was created to not publish the receiving party's public key before it is needed for verification. With this mechanism, the public key will be known when the receiving party uses the outputs by publishing a signature, including the public key. This type of transaction was in a 2018 survey found to be the most used type of transaction [BMS18]. Multi-signature (multisig) transactions allow defining the need for multiple signatures (m of n) to retrieve the funds. This allows escrow services to be built. An escrow service is required to make an escrow payment. This is a special arrangement in which the payer does not send the money directly to the payee but places it temporarily in the care of a third party [Tak17]. After the payer is satisfied that the payee has fulfilled its part of the bargain, the payer signs a transaction together with the payee. In case of a dispute, the funds are withheld by the payer. The escrow can then decide to release the funds to the payer or payee by signing the transaction with him. This is achieved with the help of multisig transactions (2 of 3). Return transactions allow a return of 83 bytes of arbitrary data. Since this data can only be removed with much effort, it has been used to store information that should not be modified, such as stamping services. The second most common transaction type in Bitcoin was in a 2018 survey 'pay to script hash' [BMS18]. It allows defining a script that needs to be known to access the funding and is, therefore, versatile.

3.3.3 Merkle Trees

The block header also includes a 'Merkle root hash' of all the transactions. A Merkle tree root hash is the root of a binary tree. The node's values of the binary hash tree are calculated by hashing the concatenation of the two descending nodes. The leaf nodes are transactions stored in that block. With this construct, a participant can prove the inclusion of a transaction in a particular block by only providing the hashes from the leaf to the root node and their neighbour values. This allows a thin client (see Section 3.2) that only has the blockchain's header information to verify if a receiving transaction really existed in a particular block (but not if it has been spent).

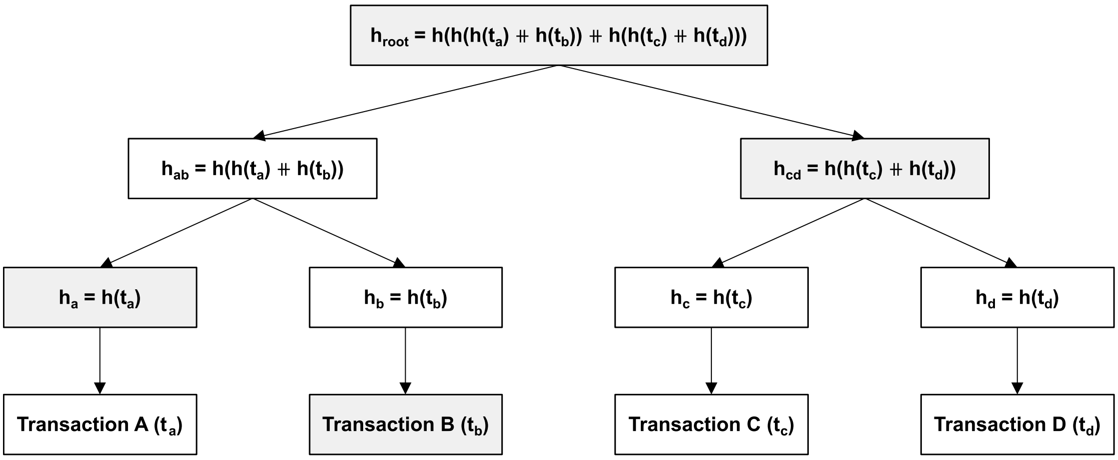


Figure 3.4: Calculation mechanism of a merle tree with transactions

Figure 3.4 shows a Merkle tree with four transactions . It shows the leave nodes of the binary tree as transactions ($\mathrm{t_a}$ to $\mathrm{t_d}$) that are hashed to nodes using the hash function h (see Equation 3.1). The parent of the hashed transactions h_{ab} is calculated by concatenating both hash values and hashing the result again. The root h_{root} is recursively calculated by the concatenation and hashing of his child nodes. When receiving a transaction to be checked if it is within a certain block, the receiving party only needs the transaction t_b, the hash of transaction a t_a and the hash of the remaining sub-tree h_{cd}. The check can be recursively calculated. When eliminating all recursions the check can be calculated as shown with function $f(h_{root}, h_{cd}, h_a, t_b)$ defined in Equation 3.2.

$$h = \mathrm{hash}(m); \tag{3.1}$$

$$f(h_{root}, h_{cd}, h_a, t_b) = \begin{cases} true, & \text{if } h_{root} = h(h(h_a \mathbin{+\!\!+} h(t_b)) \mathbin{+\!\!+} h_{cd}) \\ false, & \text{otherwise} \end{cases} \tag{3.2}$$

The size of each block and, therefore, the Merkle tree is limited by definitions. The size of each block consists of header and transaction information (see Figure 3.3) and therefore defines indirectly how many transactions can be stored in one block. The current theoretical limitation of Bitcoin blocks is about 4 Mbytes. The practical size lies at about 1.3 Mbytes [Blo21]. This is due to the fact that the block size is limited by a concept called block weight. Ethereum[3] votes on the limits in execution complexity (gas limit) and size. Since Ethereum defined a more complex scripting language that allows for loops, it needed to limit the execution of scripts. Bitcoin has a set of currently used transaction types and, therefore, a more predictable execution time per transaction type (still depending on the input counts). Ethereum implemented a Turing complete scripting language that therefore needs another form of cost calculation. Ethereum, therefore, implemented a pay-per-use model, where a transaction emitter pays a transaction fee, called gas, based on the computation complexity. This gas is also used to prioritise transactions since the emitter can pay a higher price for gas and therefore Ether (called gwei, whereas 1 gwei equals 10^{-9} Ether) for funding the transaction to incentivise the miner to execute a transaction with a higher gas price before other transactions with a lower gas price.

3.3.4 Wallets

Wallets are programs with data structures to store private and public keys (key management) and to emit transactions [SB17]. The wallet software itself is usually a simplified version of a node, also called a ‘thin client’, or an interface for remote data storage on a web server, sometimes called a ‘cloud wallet’, which is a form of custodial wallet. Custodial wallets are characterised by the fact that the private keys are held in trust by a

[3] https://ethereum.org

trustee. In mobile payment cases, wallets are rarely full nodes because of the bandwidth and storage requirements.

A wallet enables the user to send, receive and store units of cryptocurrencies in terms of the private key needed to access the fundings [BS18]. It also stores the references to transactions relevant to a user. This means that the digital wallet also serves as an interface to the network of the cryptocurrency used, for example, to store the user's balance and to display the charges for a transaction to be made. Wallets also allow for transactions for executing smart contracts. Wallets can be integrated into browser extensions in order for a website to integrate interactions with the wallet and the blockchain. This allows a website to create transactions (with the consent of the user) to, for example, log in with the wallet or buy an asset with cryptocurrencies.

Wallets are available as software for commercially available hardware such as smartphones and PCs, but also on embedded systems, i.e. on special hardware that physically secures the data, especially the private keys, on a portable storage device. The latter method is considered more secure because a hardware wallet is not constantly connected to the Internet and is, therefore, more difficult to attack, although other factors such as the computer to which the wallet is connected and the user operating the wallet play an important role in the overall security of the wallet [AGKK19]. Hardware wallets store the private keys on a secure element that does not allow extraction of the private key - not even by the owner. But it allows for signing transactions and therefore proving the ownership of an asset and spending it.

Also, wallets with integrated currency exchange features exist. These allow exchanging cryptocurrencies or local currencies for cryptocurrencies. A survey in 2017 [HR17] showed that 52% of wallets provide integrated exchange features.

Private keys of wallets need to be backed up to protect them from being lost in case of a hardware failure. Therefore, wallets usually offer the functionality to print out private keys. Alternatively, private keys can be deterministically derived from a random value that can be encoded and printed as a list of words (as defined by BIP39 [PRVB13]). Since multiple addresses should be used for privacy reasons, multiple private keys are needed. Child Key Derivation (CKD) allows generating multiple keys from one seed. BIP32 [Wui12] defines a function that generates three keys from one seed. Recursively called a tree of keys can be generated forming a Hierarchical Deterministic (HD) wallet.

Wallets are crucial for the security of funds but also for the protection of a subject's privacy. Knowing the private keys or even the seed would allow an attacker to retrace past transactions or even predict future keys and thus surveil the subject's behaviour. An informed decision on the choice of a wallet is therefore essential.

3.4 Blockchain Protocols

Based on the data structures, a communication protocol defines how this data is exchanged between the participants. A communication protocol is "a set of rules that governs the interactions of communicating entities in a communication system" [HHL94]. Each blockchain implementation has its specific protocols on different layers.

Consensus mechanisms are an important part of blockchain protocols. They define how a mutual understanding of what data is put on the blockchain is reached [SB17]. Consensus is the process for determining a valid overreaching state by all affected nodes of a network [CV17]. In the case of decentralised cryptocurrencies, this is referred to as decentralised consensus. It involves an agreement within the network that the transaction is valid and will be stored in a specific block so that the blockchain is identical again for (nearly) all participants. The consensus is, therefore, the basis for the protocol that governs the operation of a blockchain [SB17].

This section explains how these networks are built by participants based on the Internet and how the most commonly used consensus protocols work.

3.4.1 Overlay Networks

An overlay network is a logical network built on top of another network, the so-called underlay network, whereas it uses the underlay network for basic networking functions, namely routing and forwarding [Tar10]. Instead of physical machines and physical connections, an overlay network consists of nodes and logical connections in the underlying network (see Figure 3.5). It can either be a structured or unstructured network. The structure of a structured network is tightly controlled by the system in order to be able to access data efficiently [Tar10]. Overlay networks are usually built as peer-to-peer networks. Peer-to-peer networks are distributed systems where each node acts as an equally privileged client and server, meaning that they consume and serve at the same time.

The overlay network in blockchain networks is usually a mixed network with different types of nodes, formations and layers. Bitcoin, for example, has miners, full-nodes, light nodes (also called Simplified Payment Verification (SPV) nodes), and formations such as mining pools and overlays on top such as the lightning network[4]. The overlay network, therefore, is a logical abstraction of the underlying network. The overlay network is created for a use-case or application that does not need the underlying physical nodes or other communication infrastructure but for connection purposes. The lightning network mentioned above is built on top of the Bitcoin blockchain (and others) and serves as a high throughput, low latency communication channel between nodes for instant payments. This network is built on top of another blockchain network and is, therefore, also called a 'layer 2' network.

Initially, a peer-to-peer network node searches for other nodes. Seed nodes are statically programmed into the code as a starting point to connect to. These seed nodes return their list of known nodes. When connecting to these nodes, they, on the one hand, relay the information about the new node to their known nodes and, on the other hand, return

[4] https://lightning.network

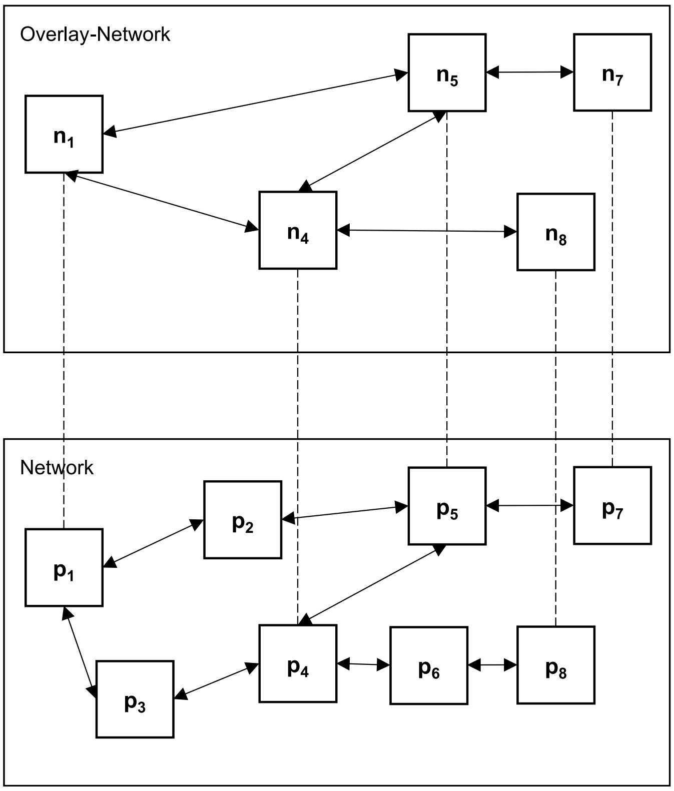


Figure 3.5: Overlay Network

parts of their list of known nodes [Ant14]. This results in direct connections between some nodes and indirect connections between (nearly) all nodes.

The resulting network is eventually synchronous, meaning the network is asynchronous but eventually gets synchronous after a while in terms of all messages being successfully delivered within the network. The two desirable properties of a blockchain network are safety and liveness. Nodes should never violate their consistency properties (safety) during asynchronous periods [CV17]. When the network stabilises and gets synchronous, then the nodes are guaranteed to run the protocol (liveness) [CV17], meaning that they validate and add transactions.

With the help of this eventually synchronous network, the nodes can exchange information and prove their entitlement to active participation, especially the generation of blocks.

3.4.2 Proof of Work

Two well-known consensus implementations are Proof of Work (PoW) and Proof of Stake (PoS). The principle of Proof of Work (PoW) is a hard-to-calculate function. The idea of ‘pricing functions’ was introduced by Dwork and Naor [DN92] to fight against junk mail. PoW is used in the mining process of blockchains as a security mechanism. It secures the chain by calculating an integral part of the chaining of the blocks and adds real costs for a potential attacker that make a successful attack unprofitable. It describes a system that requires a significant but feasible computational effort to prevent malicious interference.

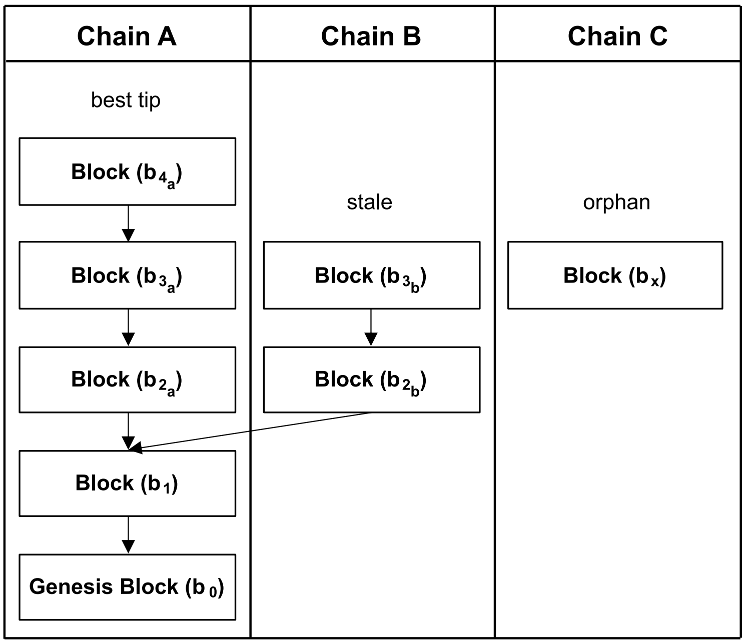


Figure 3.6: Most Cumulative Work Chain

$$[h] \sum_{n=0}^{bh=3} PoW(b_{n_b}) < \sum_{n=0}^{bh=4} PoW(b_{n_a}) \tag{3.3}$$

A blockchain network consists of many mining nodes, each of which has a specific computational power that is used to calculate hashes. The blockchain itself does not consist of only one chain of blocks (as shown in simplified figures) but merely a tree. Multiple versions of the chain compete against each other since the nodes are eventually synchronous, meaning the network is asynchronous but eventually gets synchronous over time (see Section 3.4.1). A miner starts with the first block that advances the 'longest' chain and starts the mining process based on this block ('best tip' in 3.6). If two miners find additional blocks at nearly the same time, some miners work on another 'best tip' than others, resulting in multiple competing chains.

Figure 3.6 shows how the block creation in a PoW blockchain proceeds in a way that the chain with the most cumulative work wins. The most cumulative work is defined by the most PoW done in total (see Equation 3.3). The equation uses the PoW function to determine the target difficulty of the block. It sums up the proof of work (sum of target difficulties) on both chains and compares it. In this scenario, 'Chain A' has a higher summed proof of work than 'Chain B'. It should be noted that this value is, in fact, related to the height of the chain (block height or 'bh'), but the height is not used as a criterion, as is often wrongly assumed.

Figure 3.6 shows scenarios with different chains (Chain A-C) on one blockchain. The scenario comprises two miners and four nodes. Initially, the blockchain has a length of two blocks (b_0 and b_1 in Figure 3.6). Now, two nodes (miners) are generating new blocks at almost the same time ($\mathrm{b2_a}$ and $\mathrm{b2_b}$ in Figure 3.6), expanding the blockchain, which they both are trying to propagate across the network. The miner creating 'chain A' has a faster connection to three nodes in the network and propagates its block onto them. The miner creating 'chain B' reaches only one other node in the same amount of time. This results in four nodes having version 'A' of the blockchain ('group A') and two nodes having version 'B' ('group B') at that time. The blockchain is in a forked state. This conflict is resolved after the next block is found. To determine the outcome, it is important which of the two groups finds the next block first. Since group A has a higher chance of finding a new block, since more nodes received it, its version of the blockchain remains since it produced the best tip and is propagated throughout the network, encouraging group B to discard its last block and accept group A's new block. Figure 3.6 illustrates that blocks $\mathrm{b2_b}$ and $\mathrm{b3_b}$ are no longer part of the blockchain, i.e., its miners do not get rewards for finding them, nor are the transactions in these blocks accepted; though it is likely that the transactions exist in both blocks or will be accepted at a later time. Such blocks are called 'stale blocks'. Because of that, it is recommended to wait until several blocks are added after a transaction to be sure that a certain transaction is part of the best blockchain. A block that has no predecessor is called an orphan block ($\mathrm{b_x}$ in Figure 3.6), because it does not have a known parent block. These blocks can occur when multiple blocks are found in a short period of time, and a node has not received the parent yet.

3.4.3 Proof of Stake

Proof-of-stake, or PoS for short, is used for the same purpose as PoW, namely to validate transactions. The difference, however, lies in the process. PoW requires a miner to calculate a proof that is significantly power-intense, especially since multiple miners work in parallel. The annual energy demand of Bitcoin alone has been estimated at approximately 131 TWh per year in 2021, which are 62 megatons of CO_2 equivalents [RLWK21]. The first miner to achieve this proof with a satisfying solution (over the target difficulty) is rewarded. In PoS, in contrast, it is the staking of something owned as collateral, such as a share or a token, that entitles a validator to validate a transaction and collect a transaction fee for it. In the PoS concept, anyone who owns enough shares of the currency can become a validator. Depending on the shares, validators are allowed to validate up to a certain number of blocks without having to solve computational problems. This means that if a validator, for example, stakes 4% of the units of a currency, that validator can also theoretically validate 4% of the blocks and collect their transaction fees.

There are many PoS implementations. One of the first functioning PoS cryptocurrency was Peercoin[5], first introduced in 2012 [RLWK21]. Currently there are many other PoS blockchains such as: Polkadot[6], Cardano[7] and Solana[8]. Ethereum is changing from PoW to PoS mechanism with the introduction of Ethereum 2 and is expected to "reduce carbon emissions related to the mining of ether by 99%"[Fox21], according to Vitalik Buterin, founder of Ethereum.

In this particular implementation, a validator needs to lock units of Ether (ETH), the unit of exchange on the Ethereum blockchain, to be able to validate blocks as a deposit. This process is called 'staking'. Should another node detect an attack, the propagation of the block is aborted, and, parts of the verifier's deposit are slashed (destroyed) [Sal20]. The percentage depends on how many validators are slashed at the same time ('correlation penalty'). This mechanism should incentivise a validator to validate blocks according to the predefined rules and disincentives attackers because an attack would be costly. Currently, a user needs to stake 32 ETH (about EUR 107 000[9]) to become a validator [Fou21]. Validators are randomly selected to create blocks but are also responsible for checking and confirming (attesting) all newly created blocks. In Ethereum, it is, therefore, necessary for validators to be online all the time. A downtime, and thus a non-confirmation, is penalised with a deduction of the staked Ether (liveness penalty). Selecting validators weighted by their stake is a complicated problem since there are competing interests in the network. 'Stake grinding' describes the behaviour of a validator or group of validators to manipulate the selection in their favour in order to gain (more) control over the network [But17]. It is, therefore, important to deduct the randomness from information that cannot be manipulated by users.

The Ethereum PoS implementation consists of a so-called 'beacon chain', and in future also, multiple 'shard chains' are planned. The beacon chain is responsible for managing the PoS. This includes registering the staking of validators, selecting the validators and recording attestations. The shard chains are responsible for managing the actual transactions in blocks. To process a transaction, the following procedure is followed, according to the Ethereum documentation [Fou21]:

1. a node broadcasts a transaction
2. a validator is chosen by the beacon chain to propose a new block (proposer)
3. the proposer creates a block according to the rules and broadcasts it
4. the other validators are asked to check the block and submit their attestation to the beacon chain
5. after 128 validators (committee) validate the block within a given time-frame (slot), a block is considered valid
6. an entry on the beacon chain is created
7. the proposer receives a reward

[5] https://peercoin.net
[6] https://polkadot.network
[7] https://cardano.org
[8] https://solana.com
[9] 29.12.2021 at coinmarketcap.com

As described in PoW (Section 3.4.2), there are actually multiple chains concurrently competing against each other. The difference in PoS is that the selection of the 'best tip' or 'head' is not based on most cumulative work but on a 'fork-choice rule' [BHK$^+$20]. This rule defines that the chain with the most activity weighted by stakes wins. This rule states that in the case of two parents, the one with the most attestations is chosen, with these being weighted according to the emitter's stake.

In summary, the Proof of Stake system replaces the resource-consuming Proof of Work system in order to reach a consensus on which blocks and thus transactions are included in the blockchain. For this purpose, a complex incentive-driven system is set up whose security is based on the design decisions of the operators. However, unlike the Proof of Work system, these are not those who have the most computing power through powerful mining hardware but those who have bought into the Proof of Work blockchain the most, i.e. shareholders. Thus, the incentivisation mechanism assumes that a shareholder who pays the penalty, namely his share, in the event of a violation of the defined rules, will not do so, otherwise harming herself or himself. Only time will tell whether this system has false assumptions, unknown social phenomena, or technical conceptual or implementation errors.

3.4.4 Value Transfer

There are two types of transactions, the coin base transactions and regular transactions. The former are special transactions that involve the introduction of new coins. This type of transaction is the first transaction in each block and represents the reward for solving the computational problem (mining) in the course of the Proof of Work (PoW) process. Regular simple transactions transfer value from one party to another or the same party [Oku14]. Advanced transactions can have multiple parties involved. Parties are only represented by addresses, whereas one party can generate as many addresses as it wants. Therefore the factual transfer happens between addresses, which are controlled by a receiving and sending party, meaning that the receiving party knows the private key of the receiving address, and the sending party only knows the receiving address.

A transaction is submitted by a participant of the network. This request includes necessary information for others to fully validate this transaction, such as the transaction's structure and funding. The transaction is propagated in the peer-to-peer network to eventually all participants. Each participant can validate the transaction itself. A participant called a miner puts several open transactions together in a block and calculates the block's hash. If the hash value satisfies a certain level of difficulty, it broadcasts the block as a new block of the blockchain. When successfully included in the heaviest chain (see 3.4.2), it receives the transaction fees of the block's transactions and a mining reward. A miner is, therefore, any participant that decides to create a block and calculate the hash. Since it receives the transaction fees, it will select the transactions accordingly to optimise the total value of the block in terms of size and total fees.

Figure 3.7 shows the funding of transactions via cryptography. It shows two transactions,

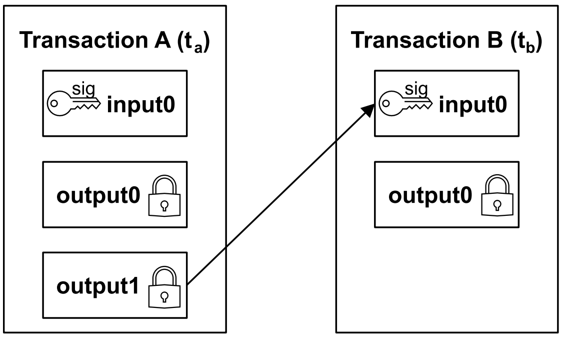


Figure 3.7: Cryptographic Funding of Transactions

whereas transaction A occurs before transaction B. It shows how transactions can transfer value from one address to another. Addresses depend on cryptographic hashes of public keys that are derived from the public key (or other mechanisms in more complicated scenarios such as a pay-to-script hash transaction). Transaction A has one input called 'input1'. This input is spent on two outputs, 'output0' and 'output1'. The 'output1' script defines that someone is only allowed to take the funding knowing the private key S of public key X, called $S(X)$. 'Input0' proves the knowledge of the secret private key $S(X)$ by providing a signature done with $S(X)$. Therefore the transaction is allowed to use that output of transaction A as funding in transaction B. The new 'output0' script in transaction B defines that whoever knows the private key S of public key Y, $S(Y)$ can take the funding. This shows the transfer from public key X to public key Y, whereas the emitter of transaction A only knows the hash of the public key X, and the emitter of transaction B knows the private key $S(X)$ and the public key Y.

Cryptocurrencies such as Bitcoin or Ethereum digitise the value on their blockchain. Whether these values are 'currencies' or 'money' is discussed in economic, legal and social aspects. Jevons [Jev83] described properties of money with utility and value, portability, indestructibility, homogeneity, divisibility, stability of value, and cognisability. The property of value given by people, no matter what the money is, is important, as is the utility as the basis for value. 'Portability' is a necessary property in order to be able to exchange money for goods or services. The property of 'indestructibility' sounds far-fetched but means that money should be durable in terms that it does not decay. 'Homogeneity', sometimes called fungibility, means that each unit of money should have the same quality and value, so it is easier to exchange. The individual units are effectively interchangeable because each of their individual parts is indistinguishable from another. 'Divisibility' is meant that not only one unit exists, but a unit can be devised in a sub-unit. The property of 'stability' of value against a utility is necessary in order to be able to have long-lasting contracts or exchanges. It is obviously an impossible superlative that is not met by any currency but should be converged. 'Cognisability' means that a currency received can be checked easily.

Figure 3.8 shows that transactions on blockchains are used to transfer value in terms

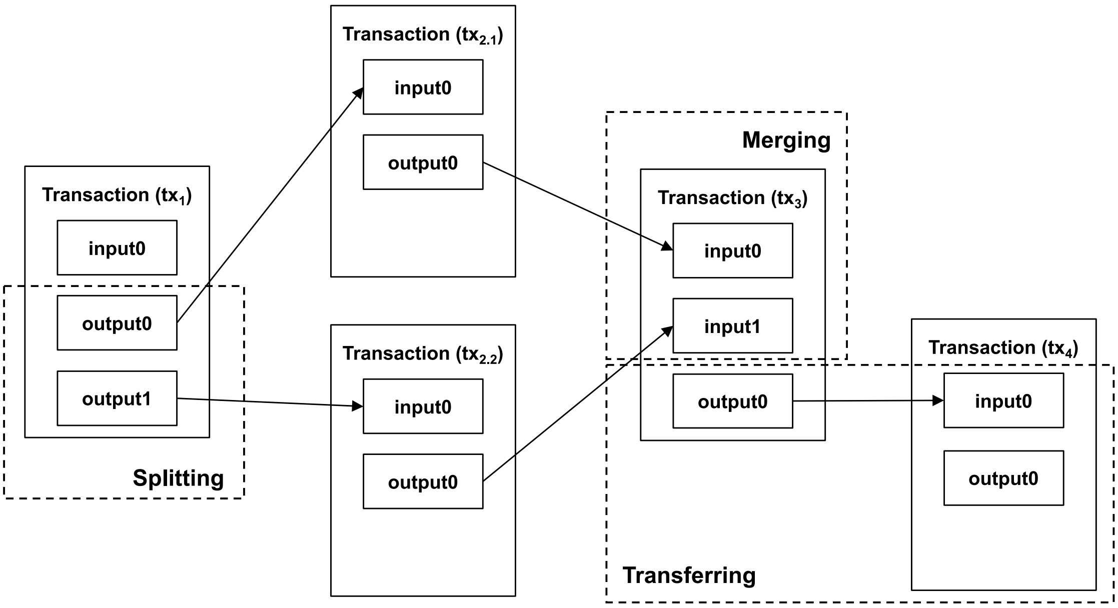


Figure 3.8: Monetary features of blockchain transactions

of monetary properties. The first quarter of the figure shows the splitting functionality or divisibility. The shown transactions receive their funding from one input transaction, ‘input0’, and splits it into two output transactions, ‘output0’ and ‘output1’. These kinds of transactions are very common since an input transaction usually does not have the exact needed value. Therefore one output is used as a return or change output, and the other output is used as the actual transfer. The second quarter shows the merging feature. Two outputs can be used as funding in one transaction. The ‘output1’ in this transaction is the sum of both inputs minus transaction fees. The third quarter shows the transfer or portability functionality. It simply transfers funding from one output to another input. These usually occur when transferring the value to another person.

3.5 Usage of Blockchain Technologies

The use of blockchain technologies in different sectors has seen a rapid gain in the past years. The International Data Corporation (IDC) has seen a growth rate of about 50% every year since 2017 and estimated a five-year Compound Annual Growth Rate (CAGR) of 46% until 2024 [Cor20]. Blockchain technologies are used in many sectors, but still mostly in banking, says the same report. The use cases of blockchain technologies in the different sectors vary. Deloitte’s blockchain survey [BBB20] asked 1488 businesses around the world whether they are working on blockchain use cases and, if so, which use cases they work on. Figure 3.9 shows the result of a survey. As expected, digital currency is the most anticipated use case for blockchain. Data access or sharing, in general, has been given as an answer by a third of the businesses. Data reconciliation, i.e. the comparison

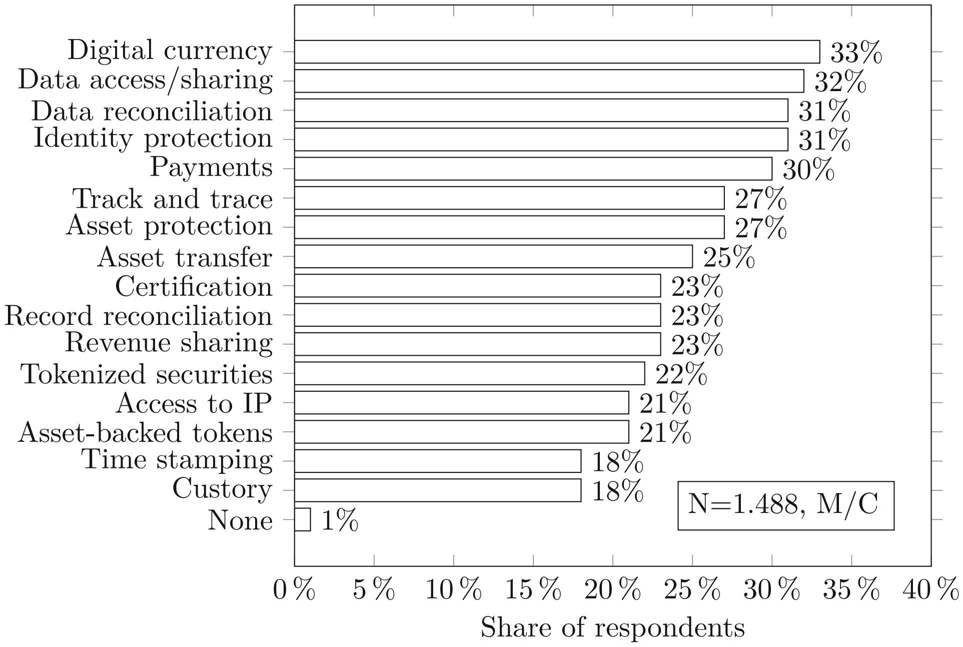


Figure 3.9: Survey of blockchain technology use cases in organisations worldwide as of 2020 [BBB20]

of data before and after a transfer or migration, is also used, as is identity management with the help of the blockchain.

3.5.1 Past and Current use of Blockchains

The banking sector was an early adopter of blockchain technology, probably in fear of the consequences of missing out on the new technologies. Namely, Ant Financial (Alipay), Visa and Mastercard together filed more than 200 patents in the European Patent Office as of 2020 [Pat20]. One of Visa's patents is related to the migration of physical currency to the blockchain by storing the serial number and value on a blockchain. Ant Financial as the owner of Alipay, spreads patents regarding use cases in payment, finance, insurance and security [Res20]. China is, in general, by far the country with the most companies applying for blockchain-related patents [inc20] [Res20].

Governmental use cases make use of the openness and availability of blockchain technologies. They mostly use blockchain for registries that are publicly available such as registration of land and property or vehicles that have been created [RVC19].

The town of Zug in Switzerland implemented a Decentralized Identifier (DID) for public identification in their town. The actual ID of the citizen is verified by the city clerk with a government-issued id and then digitally signed [Off18]. The identity service had limited use but showed how personal identity and verification processes could work with governmental IDs as a verification method. Since the 2019 Telecommunications Work

Programme of the Connecting Europe Facility (CEF) in February 2019, the European Commission has started implementing the European self-sovereign identity framework (ESSIF) as part of the European blockchain service infrastructure (EBSI). It defines besides DID and self-sovereign identity several other use-cases [Pas21]:

- **Diploma Management:** Citizens gain digital control of their educational credentials, significantly reducing verification costs and improving trust in documents' authenticity.
- **Document Traceability:** storing immutable reference data of documents or other digital artefacts that can be used at a later stage as proof of their authenticity/integrity and can be linked together to build a trusted, timestamped audit trail.
- **Trust Data Sharing:** Securely share data (such as IOSS VAT identification numbers and import one-stop-shop) among customs and tax authorities in the EU.
- **SME Financing:** Opening up new sources of (co)-finance political efforts in the area of a sustainable economy, innovation and SMEs modernisation via an EU-wide platform for debt financing.
- **European Social Security Pass (ESSP):** Prevention of fraud or error by ensuring easier communication and data exchange between European countries and the EU Institutions.
- **Asylum Process Management:** Facilitation of the management of cross-border and cross-authority processes in dealing with asylum applicants.

In the legal field, intellectual property directories that contain information about the intellectual property of immaterial goods, such as pictures, have been developed. Timestamping services or notary services for the Proof of Existence (PoE) of documents or generally files are being developed. OpenTimestamps[10] is such a service, operating free of cost. Its function is to receive hashes of files and commit them to a public blockchain, including a timestamp. To be able to provide this service for free, all hashes are put together in a Merkle tree (see Section 3.3.3), which allows adding multiple hash entries without higher transaction costs. The service is live and commits to the Bitcoin blockchain and optionally relies on calendar services, providing a faster response of validated timestamps [Tod16]. This allows to add a fingerprint of a file, such as software, e-mail, legal evidence or any other file, to the blockchain. Therefore it can be used as a feature needed for notary services, needing to prove documents existed at a given time.

IoT is a field where sensors with minimal computing power connect to a network, transfer information and execute commands, or formally defined by the International Telecommunication Union (ITU) as a "global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies" [(IT12]. [FCFL18] describes and asses the field for the use of blockchain in IoT. They find the following fields to apply blockchain technologies on:

[10] https://opentimestamps.org

- sensing
- data storage
- identity management
- timestamping services
- smart living applications
- intelligent transportation
- systems
- wearables
- supply chain management
- mobile crowdsensing
- cyber law
- security in mission-critical scenarios

In Healthcare, several applications of blockchain technology have been suggested. The health information of a patient is considered sensitive, which leads to considerable issues in the application. [EAHL16] suggest a system to share Electronic Health Records (EHRs) between providers by introducing a PoW chain handling references to records.

The automotive industry has found use cases in the field of insurance, where smart contracts predefine the terms of the insurance and pay out insurance in the case of risk manifestation [DTF19]. Also, public registries to avoid fraud have been created, such as mileage and damage databases. [PBSBS21] suggest a mileage storage based on the Ethereum blockchain with a connection to the cars on board units to track the mileage of the car. [RVC19] suggests a system for car registration based on Hyper Ledger called 'BCar'. [CIMM21] suggest a generalised data exchange system for vehicles, where vehicles, manufacturers, insurances and road maintainers can communicate about the status of the vehicle and the road.

The transportation and logistics industry has successfully used blockchain for logging transportation. Many [DHFK18, PMR18] analysed the topic of logistics and came to the conclusion that the mutual distrust of logistics processes is a good fit for blockchain.

Some of the mentioned applications of blockchain technology are theoretical concepts, others have been implemented, and others are already well-established. However, this does not mean that the application of blockchain technologies was necessary or a good idea. The next section will discuss this issue.

3.5.2 Decision to use Blockchain Technology

The use of blockchain in many sectors is a mixture of hype and necessity. Many [LXCL17, Pec17, PRB19] tried to model a decision tree to decide whether a blockchain or a private, permissioned, or public is needed. Table 3.1 summarises factors identified in these publications.

'multiple parties with shared access' means that multiple independent parties have access to the data being stored. A common database is usually only accessed by one party. It

Factor	**[LXCL17]**	**[Pec17]**	**[PRB19]**
multiple parties with shared access	✓	✓	✓
requirement for a trusted third party	✓	✓	✓
mutual distrust with conflicting interests		✓	✓
privacy of data	✓	✓	✓
write restrictions and consensus		✓	✓
censorship protection		✓	
immutable change history	✓		✓
transaction frequency	✓		
frequency of change in transaction rules			✓

Table 3.1: comparison of decision factors for the use of blockchain in the literature

also means, in contrast to data exchange systems, that data needs to be stored and that a state needs to be agreed on. All considered literature ([LXCL17, Pec17, PRB19]) had a similar factor as one of the first decisions to be made. If the data is only needed within one company, there is usually no need for a blockchain. Properties, such as immutability and audit logs, can be reached with certain databases or database extensions.

The 'requirement for a trusted third party' was also part of all considered decision trees. The formulation differs from the possibility of the use of a third party and the avoidance of a third party. The essence of the question is the idea of a blockchain to avoid a third party being trusted. A third party that is trusted by all involved parties is needed for the escalation of disputes and mutually-accepted decisions. Third parties are also involved in conventional use cases, such as central bookkeeping, escrow, notary services, and identity providers. These third parties are, to a great extent, substituted by code. The decisions and data to base the decisions on are predefined in scripts or smart contracts and executed by a party if needed. In contrast, a general trusted third party such as a state enforcing the commitments defined by a contract is not meant by this decision, even if the legal possibilities of the parties to enforce the performance of a smart contract are limited [MS19].

'Mutual distrust' is usually the reason for the need for a third party mentioned before. Two parties concluding a contract might distrust each other and therefore need a conventional ironclad contract. When using blockchain, they might only trust the code of the blockchain and the smart contracts defined. 'Conflicting interests' [PRB19] tend to lead to distrust. A flight delay insurance contract might, for example, have two interests: on the one hand, the interest by the insurance to pay as little as possible and, on the other hand, the insured to get as much as possible. This specific conflict is solved by defining the source of flight delays, a so-called oracle, and the contract terms, both in a smart contract. The interests of a party are defined by the factors leading to a payout for a certain party. If there was trust or enforceability of the contract, a blockchain might not be needed.

'Privacy' or, in contrast, 'public access' or 'transparency' of the data stored on the blockchain is mentioned by all considered papers, whereas the three considerations differ in the details. [PRB19] and [Pec17] define that if data does not need to be publicly accessible or 'kept private', a permissioned or private blockchain might be appropriate. [LXCL17] also considers the possibility of data encryption if 'transparency' is required. This shows the ambiguousness of the term 'transparency'. Blockchain has transaction transparency when being a public blockchain. Everyone who desires can watch the transactions being made onchain. This differs from data transparency, where everyone can see and interpret the data. Whether encryption is a suitable mechanism is discussed in Section 4.2.3. The fact that all three papers mention it as an integral decision whether to use a blockchain shows the importance of the topic being discussed more thoroughly.

In addition to read-restrictions [Pec17, PRB19] discuss the need for write and update restrictions and consent over the correctness. Whereas [Pec17] discusses the need for any change done in general by multiple parties [PRB19] discuss the need for consensus. As described in Section 3.4, consensus needs to be reached on whether an update to the blockchain state via the acceptance of a transaction is accepted. The special role of a consensus node in regards to writing restrictions is, therefore, factually limited to selecting and validating transactions to predefined rules. Write restrictions in terms of access control layer are, on the one hand, implemented in special blockchains called permissioned blockchains or in extensible blockchains as smart contracts [CKY18].

The decision of readable and writeable data extends to censorship. [Pec17] highlight that a blockchain is a good tool for avoiding censorship and denial of service attacks. The reason lies, on the one hand, in the distributed nature (see Section 3.1.1) of a blockchain and distributed ledger in general and, on the other hand, the change history. [PRB19, LXCL17] highlight the immutability and changelogs. Immutability, as an overstatement of the near impossibility of change, makes censorship harder since data is hard to alter without censoring all data on the blockchain. The changelog or history helps detect censorship attempts in the final state and, if openly accessible, makes them impossible. [LXCL17] add the need for immutability, in the sense that past data must be modifiable as a decision against the blockchain, and the possibility of storing data outside the blockchain in a separate database that references the data in the blockchain. Further mechanisms to provide secure mutability are discussed in Section 4.2.3.

The performance of different blockchain implementations has been discussed in many papers [HLD$^+$18, SHN$^+$18, DLZ$^+$18, PST17]. These discussion reach from the storage needs to the number of transactions per second that can be handled by a certain blockchain. [LXCL17] decided to add the need for high transaction frequency and high volume to their decisions against the blockchain. The fact that blockchains try to solve both issues by technologies such as segwit, lightning, pruning and others shows that, on the one hand, it is an issue, but on the other hand, it shows that this issue is solvable.

Many decisions in the decision trees are absolute. The decision by [PRB19] to not use blockchain technology if there is no need for an immutable log is an example. Also, the possibilities of selecting blockchains with specific functionalities, such as encryption, off-chain storage etc., make a decision solely based on a binary decision tree hard. The complexity in such decisions lies in the fact that there is no 'standard blockchain'. It cannot be decided if there is enough throughput since there are blockchain-related technologies trying to mitigate that. It can also not be decided if a traditional database is sufficient, as suggested. A more detailed and requirements-based qualitative approach must be used to decide whether a specific blockchain can be used for a use case.

3.6 Discussion

Blockchain technology has developed from enthusiastic hype to a recognised and widespread technology [BBB20]. It is essential to classify the ambitious developments in the fields of engineering and science and to analyse them scientifically. In addition to the technology and possible applications, there are also changes in data protection and data security. These must be analysed scientifically and impartially. With the help of the presented basics, the question can now be asked whether these applications have an influence on privacy. On the one hand, the unbiased question is whether the technologies have a negative impact on data protection and what solutions to this negative impact might look like; on the other hand, the question is what privacy-supporting applications might look like and whether these, in turn, cause data protection problems. These two topics will be addressed in the following chapters.

CHAPTER 4

Data-Protected use of Blockchain

Data protection and privacy have been an issue of blockchain technology since the first mention by Satoshi Nakamoto in the famous Bitcoin whitepaper [Nak08]. He describes the notion that Bitcoin would be private without restricting access to information but by withholding the identity of the transaction or private key owner from the public. He also mentions the possibility of changing the address for each transaction against de-anonymisation but notes the possibility of tracing back a common owner by linking transactions. A simple de-anonymisation scenario (see Figure 4.1) tracks an illegal purchase (tx1) by following the (assumed) change returned by the transaction until a transaction has been made to a known entity (tx2), such as an exchange or payment provider that pays out by SEPA (payout funding). A court can then order this entity to disclose the recipient's identification information.

Even if there is no known entity involved, researchers have shown that Bitcoin addresses can be mapped to IP addresses with a reasonably high probability [RH12, TS16]. Other publications [FKP15, BKP14, MPJ$^+$13, RS13] showed different de-anonymisation techniques that can be applied to Bitcoin, all resulting in a certain probability of mapping transactions to a wallet and therefore identifiable natural person. These privacy issues are not limited to Bitcoin but also apply to other blockchain implementations such as Ethereum under certain circumstances [Tik18].

A quick, unsatisfying answer is offered by the use of privacy-enhanced blockchains. These supposedly manage to solve the privacy problems with technology. Monero[1], as such a privacy-enhanced blockchain implementation, is unfortunately also not fully protected against privacy attacks. Monero tries to enhance privacy with the use of 'CryptoNote' that allows users to obfuscate their transaction graph by inserting chaff coins, known as 'mixins', alongside the actual coins they spend by using ring signatures [VS13]. These signatures allow for multiple signatures to sign a transaction without

[1] https://getmonero.org

knowing which signature was the needed signature to authorise the transaction. This makes Monero harder to track and, therefore, less vulnerable to some deanonymisation or re-identification attacks. Older implementations of Monero were vulnerable to tracing due to poor mixin selection that allowed partial deanonymisation [MMLN17]. Monero also uses zero-knowledge proofs, which is a proof system in which a prover can prove to a verifier that he or she knows something without revealing any information beyond what the verifier already knows [VS13]. Zero-knowledge proofs are possible because of mathematical insights into the nature of truth and proof. In general, it is impossible to create a perfect proof system in which all truths can be proven and all false statements exposed. However, some proofs can provide convincing evidence without revealing anything beyond what is already known by the parties involved. This property makes them ideal for use in sensitive situations where privacy is desired. Section 4.2.3 discusses zero-knowledge proofs in detail.

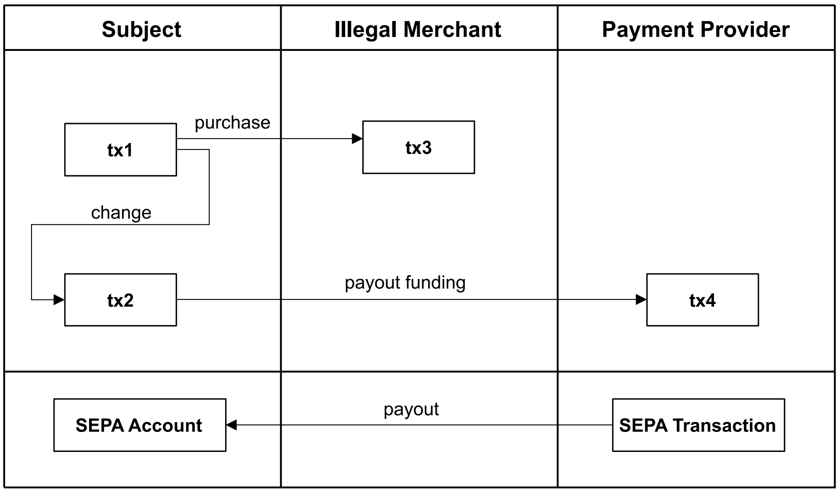


Figure 4.1: A simple de-anonymisation attack scenario by back-tracking transactions

These privacy issues and possible solutions need to be addressed. Data protection in terms of the GDPR is much broader than data security, and the attacks on anonymisation or re-identification are solved by privacy-enhancing blockchains. Properties of blockchain technologies such as decentralised authority, distributed data and immutability are desirable in many applications, as shown in Section 3.5. However, data-protected use of blockchain technologies is difficult to achieve, as these properties counteract privacy properties and values, as shown in Section 2.1. Focusing on the legal framework of the GDPR (Section 2.3), the research questions defined in Chapter 1 therefore are accordingly “How does the use of a public or private blockchain change GDPR compliance in an application?” (Question 1.1) and “What are feasible solutions for data protection issues on blockchains?” (Question 1.2).

The research questions will be answered by a theoretical analysis of the technology regarding the requirements of the GDPR with regard to data protection. A literature research regarding solutions for the found problems leads to a discussion and a result showing the supporting and obstructing features in relation to obligations of controllers and rights of data subjects answering the research questions.

4.1 Risk-based Assessment of Personal Data Processing Operations

Section 2.3.6 discussed the need for adequate measures regarding data protection according to the GDPR. Recital 84 answers the question of what approach is appropriate to select measures. It suggests assessing the risk to the rights and freedoms of natural persons to enhance compliance as part of a data protection impact assessment. This is due to the need for measures to be proportional to the risk. The following section, therefore, proposes an approach for quantifying the risk and applies it to the case of described data on the blockchain.

Risk-based data assessments are currently used in organisational contexts by identifying and assessing the risks associated with different types of data to develop more effective security plans and protocols.

One key factor in a risk-based assessment is understanding the value and the risks of the data processed. This may vary depending on the type of data, its intended use, and its sensitivity. Also, the assessment may be subjective since the mentioned attributes vary depending on the point of view. For example, identification information is valuable to criminals but may not be as important to a controller. Personal information rather is important for the individual.

It is therefore important to use an assessment process complying with the GDPR, that is suitable to analyse a technology and its underlying data processings in its data protection enhancing or obstructing features.

The following sections use the Data Protection Working Party (WP248) guidelines and the ISO 27005 [ISO13b] defined a risk assessment process to evaluate the risks originating from the use of public or private blockchains.

4.1.1 Risk-Based Data Protection Impact Assessment Process Design

Blockchain as a technology is rather difficult to assess since it summarises several systems and implementations. These themselves cannot be assessed whether they comply with the GDPR since they are processings-agnostic. Therefore different use-cases and implementations are to be analysed separately so that conclusions regarding their data protection properties can be drawn. The basis of blockchain technology, as described in Section 3.2, are data structures and protocols.

An assessment of the data processed can be done via information classification. A common method of classifying information is described in the ISO 27001 [ISO13a] standard for information security. This information classification is used to protect the information by applying the proper security controls.

ISO 27001 describes the control objective A8.2 ’Information Classification‘, which defines that data is first gathered and registered; then classified; the data being labelled according to the classification and acted upon defined rules for the classifications [ISO13a]. Organisations should take a risk-based approach to classify information, considering how likely an incident could occur and how much damage it could cause. Once the information has been classified, it must be handled and stored in a manner that ensures its security.

An organisational approach cannot be used in the case of a public blockchain since it is not operated by a single entity (see Section 4.2.1), and the impact on the rights and freedoms of the data subjects should be assessed.

[Par17a] suggest that the Data Protection Impact Assessment (DPIA) should use the following process, based on ISO 27005 citing the Recital 90:

- Establishing the context: “taking into account the nature, scope, context and purposes of the processing and the sources of the risk”;
- Assessing the risks: “assess the particular likelihood and severity of the high risk”;
- Treating the risks: “mitigating that risk” and “ensuring the protection of personal data”, and “demonstrating compliance with this Regulation”.

This fundamental process maps the Recital 90 to a standard risk management process, whereas it misses the assessment of the residual risk (risk acceptance) to determine whether the risk is still high [Cou16a, Art. 36(1)].

Figure 4.2 shows the proposed risk assessment process based on the ISO 27005 [ISO13b] and the Article 29 Data Protection Working Party (WP 248) Guidelines on Data Protection Impact Assessment (DPIA) [Par17a] to assess data in an impact assessment, implement a treatment with a feedback loop to re-assess remaining residual risks.

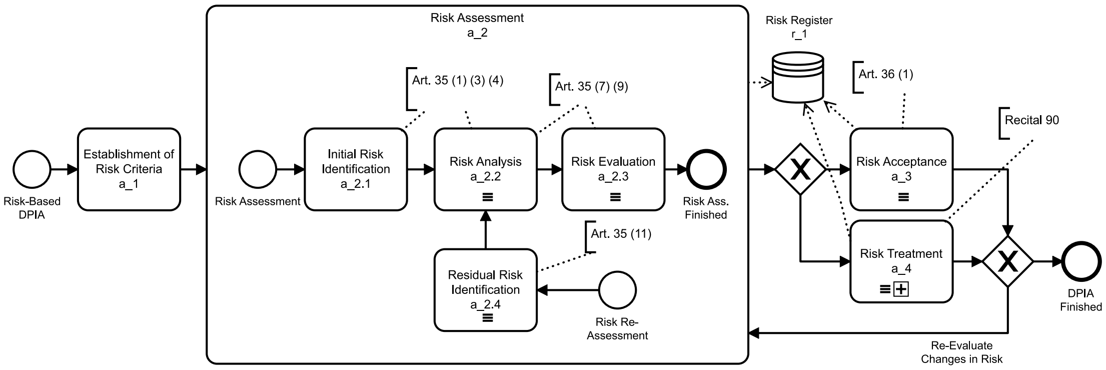


Figure 4.2: Risk-Based Assessment Process (based on ISO 27005 [ISO13b] and the WP 248 Guidelines regarding DPIA [Par17a])

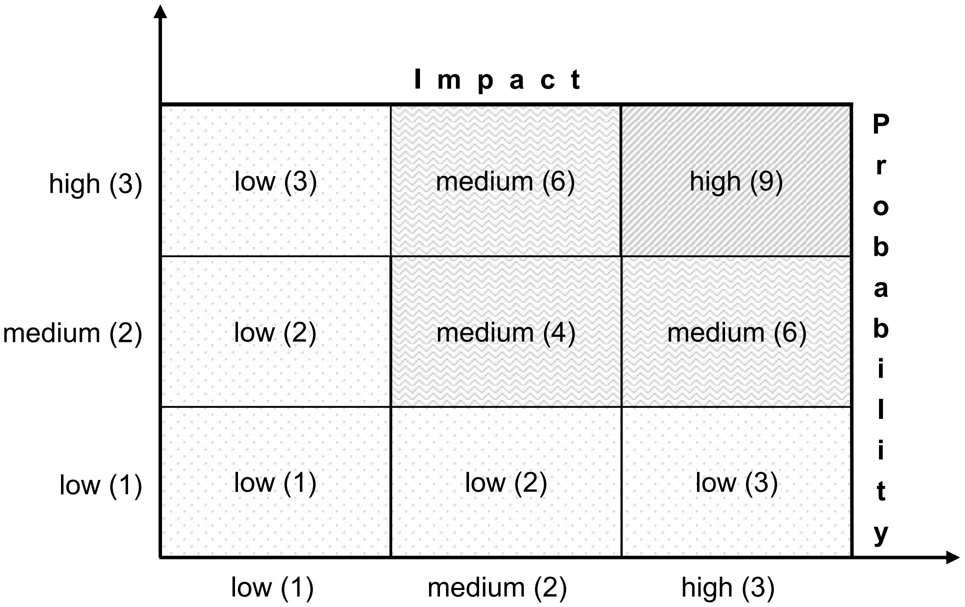
The first activity is the 'Establishment of Risk Criteria' (*a_1*). It is used to find processing activities and understand the context thereof. This activity should discover what data is processed, where it is processed, the legal basis and the overall objectives of the processing activities. Also, the risk criteria need to be defined. Risk criteria are the "terms of reference against which the significance of a risk is evaluated" [ISO13b]. Art. 32 (1) asks to consider the "state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons" [Cou16a, Art. 32 (1)]. Therefore, the criterion needs to be focused on the freedoms of natural persons, not the criticality regarding the organisation or its operational goals.

With this information at hand, the actual risk assessment (*a_2*) can start. Art. 35 (7) and (9) contains specifics regarding what the assessment must contain and that the view of the data subject must be considered. 'Initial Risk Identification' (*a_2.1*) searches for risks within the given context (processing activities, systems, etc.). [Cou16a, Art. 35 (3) and (4)] defines specifics regarding which processing activities require a data protection impact assessment. In order to envision rights and obligations in terms of the GDPR, the author suggests a risk register based on these. Table A.1 in Appendix A shows a template that allows for risk registration based on the rights and obligations defined in the GDPR. Risk registers are in general documents or databases that list risks and their associated proactive treatments and reactive responses [Ins21]. This is needed due to the requirement to demonstrate compliance with needed technical or organisational measures [Cou16a, Art. 5 (2)] and the execution of a DPIA [Cou16a, Art. 35] which might be needed to submit to the DPA [Cou16a, Rec. 94]. The proposed risk register allows traceability in a way that legal requirements, risk assessments, and risk treatments or risk acceptances are linked so that they can be traced in retrospect.

The 'Risk Analysis' (*a_2.2*) assesses the impact (low, medium, high) of each risk on the data subject and the likelihood of it happening (low, medium, high). This helps to prioritise the risks and decide which need the most attention. These are documented in the risk register (Table A.1 in Appendix A) and include both internal and external risks. Internal risks are those that are specific to the organisation, such as a data breach or unintentional destruction of data. External risks are those that are outside the organisation, such as legislation or technological advances. Art. 35 (1) asks for "prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data" [Cou16a, Art. 35 (1)]. Within this activity, it is advised to involve a data protection officer [Cou16a, Art. 35 (2)].

With the analysis at hand, the 'Risk Evaluation' (*a_2.3*) compares the results with risk criteria (defined in *a_1*) to determine whether the risk is acceptable or needs to be treated. The risk register is updated with the multiplication of the impact and probability (see Figure 4.3). The result is then linearly ranked from low to high.

If the risk can be accepted (due to the magnitude), the 'Risk Acceptance' (*a_3*) activity records the decision in the risk register. Otherwise, the 'Risk Treatment' (*a_4*) activity



Probability \ Impact	low (1)	medium (2)	high (3)
high (3)	low (3)	medium (6)	high (9)
medium (2)	low (2)	medium (4)	medium (6)
low (1)	low (1)	low (2)	low (3)

Figure 4.3: Calculation of the risk priority based on impact and probability

implements a proactive treatment. The proactive treatment is not to be confused with a reactive response. Treatments include avoidance, mitigation or transfer.

'Avoidance' tries to eliminate the threat or to protect from impacts [Ins21]. This can be implemented by changing the processing operation, deciding not to process the according data at all, or anonymising the according data. In this case, the outcome must be that there is no processing operation according to GDPR Art. 4 (2).

'Mitigation' tries to reduce the probability of the threat happening or reduce the impact [Ins21]. This can be done by technical and organisational measures such as pseudonymisation (see Section 2.3.6).

'Transfer' tries to move the threat to a third party to manage the risk or to take the eventual impact [Ins21]. This can be partially done by contracting a processor (since the controller is still accountable, see Section 2.3.3).

After the implementation of the risk treatment, a feedback loop is executed to check on the remaining/residual risks with the 'Residual Risk Identification Risk Register' (*a_2.4*) activity suggested by Art. 36 (1) and required by Art. 35 (11).

A crucial activity within the proposed process lies in the analysis of the given risk. Controllers and processors need to be able to understand the risk from the view of the data subject [Cou16a, Art. 35 (9)]. The following Section, therefore, proposes dimensions for analysing and estimating the risk that is inherent in data retention.

4.1.2 Analysing Data-originating Risks

The classification of personal data, as described in Section 2.1.3, is relatively vague since the relatability to a natural person must be given or foreseeable with technical advances. This is especially the case if relatability relies on protection mechanisms that cannot be changed over time.

The technical literature discusses whether certain data is directly personal, indirect personal, pseudonymous, de-identified, or anonymous (see Section 2.1.3). The distinction thereof is often unclear because of theoretical or not yet realisable technical possibilities. Therefore, the field of data protection by design asks if currently anonymous or pseudonymous data might be identifiable in the future. This challenges the design of such systems with an unknown factor decided in the future.

Legal discussion in this regard adds the possibility of adding data (legally) to the dataset, meaning that one party might have the legal authority to add information that makes it more identifiable (e.g. [Cou16b]).

On the legal side, discussions about the hypothetical de-anonymisation, in an absolute manner, meaning taking all existing data into account, and objective, meaning data might be classified in terms of identifiability from a data holder perspective.

Legal discussions about technical advances are already part of the legislation, demanding an objective, risk-based approach that takes technical advancements into account [FP20].

The GDPR Recital 26 asks to protect against ”reasonable likely“ attacks considering “all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments”[Cou16a, Recital 26].

In big-data applications, techniques to find groups in anonymous data collections with comparable properties exist. Technical implementations to relate previously unrelated data by analysing patterns and clustering to find useful patterns and groups is a declared objective of data mining with several approaches for clustering users to predict future behaviour [SAWH14]. The European Data Protection Supervisor therefore goes even further by stating that “big data should be considered personal even where anonymisation techniques have been applied: it is becoming and will be ever easier to infer a person’s identity by combining allegedly ‘anonymous’ data with publicly available information”[Eur15]. They, therefore, state that the “challenges and risks of big data [...] call for more effective data protection”[Eur15].

The independent European advisory body on data protection and privacy (Article 29 data protection working party) published a classification for determining if a processing operation is carried out on ‘large scale’, because it determines the need for a Data Protection Officer, mentioning the following factors [Par17b]:

- The amount of data subjects concerned
- The volume of data and/or the range of different data items being processed

- The duration, or permanence, of the data processing operation
- The geographical extent of the processing operation

Based on the Article 29 data protection working party criteria for determining large scale processing activities [Par17b] and technical properties (defined in Section 2.1.3) the author suggests a three-dimensional assessment of processed data (see Figure 4.4). These three dimensions include technical and legal perspectives in a joint representation.

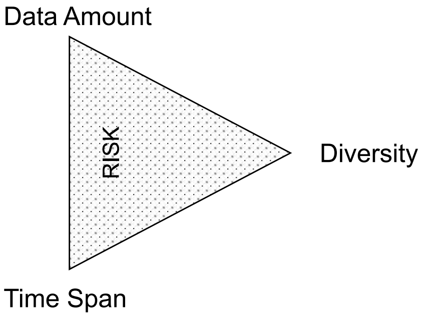


Figure 4.4: Dimensions of personal data

- **Data amount:** Horizontal and vertical amount of data, directly or indirectly related to a data subject.
- **Diversity:** Strength or relation to the data subject in terms of personability and, in contrast, anonymity (k-anonymity and l-diversity).
- **Time span:** Duration of data handling in given quantity and given reference.

'Data amount' relates to the data existing, relatable to a person. It includes horizontal as well as vertical data. The dataset under investigation must be a closed set of data (subjectively), but the data amount must be the data existing to relate to (objectively). The data amount is not to be confused with quality. Quality can define the properties of one data field in terms of desired information retrieval. The amount of data does not consider information or information density.

'Diversity' or contrary 'Relatability' describes the strength with which the dataset under investigation is related to an individual as a qualitative attribute. Personal data has either a direct or indirect connection to the subject, meaning it needs further data to be directly attributable or has a fuzzy relationship in terms of probability. Differentiating between the intrinsic diversity of data and extrinsic connectives is important. A technical measurement for diversity is l-diversity (see Section 2.1.3). Diversity can be seen as one aspect of data quality in terms of information content. In comparison, extrinsic connectiveness assumes additional knowledge and information from other data sources. Diversity is dependent on the amount of data. Also, the time span can have an influence on the relatability and actuality of data.

'Time span' relates to the time the dataset under investigation will be existing in the given quality and quantity. One can assume technical advancements, for example, in data analysis and encryption and fields not directly related to data, such as data extraction, image recognition, and classification of data. Data that has been anonymous or pseudonymous at one time might not be anonymous at a later time. Also, data is ageing, meaning that data at a current time might be highly attributable but later is less attributable. In addition, the amount and risk of data changes over time in terms of legal authority over the data. Data collected might not pose a risk at the time of collection, but the authority over the data might change. Therefore, the time data is kept is crucial when assessing risks. Also, the time span can be infinite if the data cannot be erased or destroyed. The 'given quality' mentioned above also refers to the technical possibility of data recovery, i.e. the recovery of previously deleted or destroyed data. It also refers to the 'age' or 'timeliness' of the data. Outdated data might not be as qualitative as new data.

The risk assessment with personal data has the same issue as any risk assessment. Data is only reliable to a certain degree, and circumstances that influence the manifestation of the risk are unknown or uncertain. The objective amount of data that can relate to a person cannot be known by an assessor. Data might not be available or accessible at the time of assessment. Also, published data, for example, with free access by anyone, can be copied by others, not to mention illegal access and copies of data. However, the secure deletion of data has uncertainties. The time span of data existing in a certain quality, such as encrypted data, is hard to estimate and predict — the diversity of data changes with both time and amount of data available. A dataset might be attributable to other data in hand or the right technology or computing power (e.g. hashing or encryption).

4.2 Application of Risk-Based Data Protection Impact Assessment on Blockchain Operations

The research questions defined in Chapter 1 are answered by executing the assessment process defined in Section 4.1.

First, the background and roles of the processing activities are analysed in order to gain insights into the environment. Also, the applicable criteria defined in the GDPR are listed and brought into connection with the subject matter. The risks regarding the processing activities are identified and analysed. After Evaluating these risks, potential risk treatments are discussed, and conclusions are drawn.

4.2.1 Establishment of Risk Criteria

The technical environment of blockchain has been described in Section 3.2. The GDPR defines roles with different rights and obligation (see Section 2.3.1). Identifying these roles in a processing operation is important for processors and controllers in order to know their obligations to be able to fulfil them accordingly. The aforementioned technical roles in a blockchain network, on the other hand, are technical in the sense of functional components that interface with other components and are usually operated by separate legal entities. Technically, multiple roles can be assumed by multiple entities, and one entity can assume multiple roles.

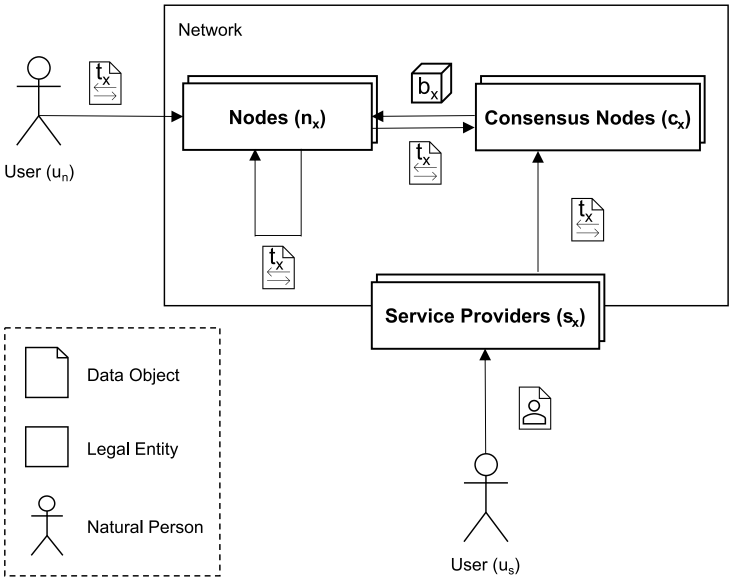


Figure 4.5: Schematic representation of GDPR roles in a blockchain environment

Figure 4.5 shows a simple model of a blockchain network, where nodes (n_x) produce transactions (t_x) and transmit them to other nodes, including consensus nodes. Consensus nodes (c_x) are nodes that reach a consensus and create blocks (b_x) with a set of transactions in them. These consensus nodes are usually miners or stakers, depending on the proof (see Section 3.4). The created blocks are again transmitted to other nodes. Service providers (s_x) provide services regarding the blockchain network to subjects not directly connected to the blockchain network. Service providers usually know the identity of their customers (data subjects) because of direct connection (IP addresses) or legal obligations (KYC obligations). Service providers might operate as a node or even consensus nodes besides providing services. Such service providers span from payment providers to simple money exchanges to full-blow multi-currency stock-marked exchanges with options and lending. Also, cloud wallets might operate as service providers with the same properties.

Whether a technical role of the network depicted in Figure 4.5 is a data processor, a controller, or a joint-controller needs to be analysed regarding its control and decisions over the processing since the controller decides “the purposes and means of processing” [Cou16a, Art. 4 (7)], whereas the processor is usually a third party that provides data processing services to the controller with no influence on the purpose and little influence on the means.

In the case of nodes (n_x), the author concluded in [SFN$^+$18] that a node “decides if a transaction is valid and to whom it is propagated” [SFN$^+$18] and will therefore most likely be seen as a controller or joint controller together with all nodes. [And20, p. 114] argue similarly that nodes' decision is freely given and not bound by directives of third parties and come to the same conclusion that nodes are either controllers or joint controllers. [PV19, p. 172] argue that nodes only work on given rules and do not decide the means independently. Furthermore, they argue that a single node's contribution is too little to be considered a controller. The European Parliament argues in its study [Fin19] that nodes provide a similar infrastructure as the payment infrastructure provided by the Society for Worldwide Interbank Financial Telecommunication (SWIFT), which has been considered a controller by the Article 29 Working Party, because of its 'significant autonomy' over the processing operations. They further argue that nodes together decide the version of the protocol and could hard-fork (see Section 3.4) the blockchain and might, therefore, together have autonomy and enough influence to be considered joint controllers.

In the case of consensus nodes, especially miners (c_x), the author argued in [SFN$^+$18] that they receive rewards and receive the data from simple nodes and might be considered processors since they do not decide the means and purpose (see Section 2.3.1). [And20, p. 112] argue that this dependency is not given. They argue equally that the processing activities are done on behalf of a third party but that they freely decide to verify transactions and do not receive any instructions from third parties. Therefore, they conclude that miners are controllers or joint controllers if they operate together. They also conclude that consensus nodes operated in private chains (usually stakers) are working on behalf of an 'initiator' or network operator and are, therefore data processors. [PV19, p. 173], on the contrary, argue that miners do not decide which transactions are to be put in a block but instead execute the programmers' rules that they must adhere to; Otherwise, their blocks would be disregarded by nodes. [Cza17] describes that the consensus mechanism run by the miner itself decides the means of processing in the blockchain network and therefore concludes that "we can potentially treat transaction-confirming miners as controllers" [Cza17]. The French national data protection authority (CNIL) [dledL18] and the EPRS [Fin19] came to the same conclusion that the means of processing are defined by the consensus nodes, but that the purposes, which overrule the means (see Section 2.3.1), are not influenced by the consensus nodes, and therefore they are not controllers.

Summarised, the issue regarding the decision of whether a consensus node is a controller lies in the question of whether it is deciding the purposes and secondary the means of the processing activities ('primacy of the purposes criterion' [Fin19]). Technically each of the aforementioned arguments is valid. A node or the operator of a node decides which transactions are executed within the block it generates; it receives a reward from the party wanting the transaction to be executed; it therefore usually chooses the transactions with the highest rewards; it has limited decision-making power of what is executed since it must adhere to a pre-defined protocol; it merely configures a program created by a third party (programmers).

The rulings on the General Data Protection Regulation so far go in a direction where joint controllership is more often assumed than not to ensure the "effective and complete protection of data subjects" [Fin19]. The programmer of an application executed by a controller or processor does, unfortunately, play a very limited role. The operator of a node can configure a node in a way that it decides the means. A Bitcoin core node, for example, can be configured to only relay transactions with a certain minimum transaction fee (*minrelaytxfee*), to not process data carrier (*OP_RETURN*) transactions, set a minimum and maximum block size, or define how many blocks are stored before they are deleted or pruned (data amount) [Lop22]. Whether this constitutes "essential elements of the means" [Par10] could be discussed. The strongest argument for controllership lies in the aforementioned configuration options. These let the operator decide the purposes of a consensus node. It can be operated in a way that allows distribution and storage of data (data carrier) or only executes other transaction types. It decides if free transactions are possible. This is not done in an ultimate way but only for the operations of its own node. If another node decides to process free or data carrier transactions, it might do so, resulting in these transactions being processed eventually. This constitutes a joint-controllership, less in a consensus before the processing activities, but by configuring processing decisions in the nodes. Therefore a joint-controllership seems to be appropriate to be assumed for consensus nodes.

In the case of service providers [And20, p. 116] argue that they usually decide the means of processing, as banks and payment processors do, and are, therefore, generally controllers. They also mention that service providers could be acting on behalf of participants and might therefore be processors. [PV19, p. 174] see service providers as controllers since the users of such services transfer the control over to the service provider, often not even controlling their private keys. Therefore, the service providers define the means of processing and are generally qualified as controllers [PV19, p. 174]. [Pei20] sees any platform 'outside' the network as its own controller, not being part of the problem of responsibilities in the blockchain system.

In the case of users, the EPRS [Fin19] follows the report of the European Parliament on the future of blockchain [McC18] that states "users may be both data controllers, for the personal data that they upload onto the ledger, and data processors, by virtue of storing a full copy of the ledger on their own computer" [McC18]. The French data protection authority has a similar tension but discusses if users might fall under the household exemption. This exemption would make the processing of users not fall under the GDPR. For this to apply, the purpose of the transactions would have to be non-professional and non-commercial.

[Fas17] argues completely differently. He describes the blockchain rules as set by the developers of nodes and miners, and platforms and therefore sees the entire responsibility with the developers. [Pei20] argues that the program code is openly accessible and changed within the behaviour of the system would need to be accepted by the network. He also argues that the GDPR Recital 78 explicitly states that developers of systems are not responsible. The EPRS publication on blockchain and the GDPR comes to the conclusion that even though developers influence the means of processing by providing an algorithm, they "are the least likely to qualify as controllers" [Fin19] since they do not pursue their own purposes, but merely make infrastructure possible for others.

[Isl17] argues that the system of the GDPR is only designed for the case of one responsible legal entity (controller) and one executing legal entity (processor), whereas blockchain needs a shared responsibility. The solution of joint responsibility as defined by the GDPR is, in his opinion, not feasible since the nodes are not a communicating, coordinated collective that is collaborating for a joint aim. [Syd18] states that according to the GDPR, there must be a controller for every processing operation. The European Parliament Research Service (EPRS) goes further by inferring from the opinion of the Advocate General at the European Court of Justice that "anyone that chooses a particular technical infrastructure, such as DLT, to process data, can be a joint-controller of that system even though they may only have limited control over the purposes and no meaningful control about the means of processing" [Fin19].

The blockchain environment is also defined by the distributed nature of the system and data, meaning that nodes and, therefore, legal entities are located in different locations. Current estimations by a project screening the Bitcoin network (Bitnode[2]) estimate that 12% of the network nodes reside in the United States, 12% in Germany, 4% in France, 3% in the Netherlands and 2% in Canada. The rest is spread around the world, and more than 50% are unknown due to the use of the onion network (Tor Project[3]). This is probably due to the fact that countries are starting to ban cryptocurrencies. China first prohibited financial institutions from engaging in any cryptocurrency transactions, then banned cryptocurrency mining, and finally prohibited cryptocurrencies completely [QG22]. Bitnode's historical data shows that before 2020 only 2% of the nodes were hidden with Tor services.

This influences the GDPR applicability and risks related to the GDPR. On the one hand, the GDPR is not applicable to most of the network. On the other hand, it can be said that a transfer to third countries happens when storing data on this particular blockchain. It can also be assumed that the needed level of protection of natural persons [Cou16a, Art. 44] is not ensured in these third countries.

The types of data processed in an application are highly application-specific, not technology-specific. Nevertheless, a technical distinction can be made between protocol data such as metadata and headers/trailers and the payload needed for the specific application. Both need to be considered in the structured identification process.

The datasets in both protocol and payload data is usually a mixed dataset. Mixed sets contain personal and non-personal data. The guidance on the Regulation on a framework for the free flow of non-personal data in the European Union [Eur19b] clarifies the use of mixed datasets and the implications regarding the GDPR. Where data can be split into personal and non-personal data only needs to be handled according to the GDPR. In the case of mixed datasets that cannot be split, rights and obligations regarding the

[2] https://bitnodes.io/nodes/#network-snapshot Accessed: 27.01.2022

[3] https://www.torproject.org

GDPR apply to the entire dataset without any consideration of the proportion of the data [Eur19b].

Therefore, in the case of blockchain technology, where datasets or blocks are immutably (see Section 3.3) linked, data cannot be split or separated. The following section discusses the roles and the data being processed by them in detail.

A 2018 study by the author already showed major concerns about both protocol and payload data on the Bitcoin and Ethereum blockchain [SFN$^+$18]. It showed that personal data could be found within transactions of Bitcoin, namely 'pay to public key' and 'pay to public key hash'. It also discusses the general intention of payment transactions as a method of transferring funds from one party to another, which intentionally relates to parties by their public keys. The receiving party discloses its public key at the latest when using the received funds and therefore adds personal data to the blockchain. Latest after the clarification in Recital 26 [Cou16a, Rec. 26] regarding hashing as a pseudonymisation instrument, it can be assumed that even before using the funds, the emitter discloses personal data of the receiving party.

Anderl and Schelling argue in [And20, p. 99] that the same relationship between a public key and a natural person exists, as in the relation between IP address and natural person. Since the case of IP addresses has been famously decided in [Cou16b], they argue that public keys are generally to be qualified as personal data [And20, p. 99]. Even though they come to the same conclusion, the argument has a critical failure. They assume that public keys have an awarding authority that knows the natural person's identity and can therefore re-identify the natural person, comparing it to the Internet Service Provider (ISP) handing out IP addresses. This is, in general, true for public keys in a Public Key Infrastructure (PKI), but is not used in the leading public blockchains, Ethereum and Bitcoin.

Since public keys are uniquely assigned to individuals, they could also be re-identified via external data sources. Service providers such as Internet service providers but also exchanges require identification by means of a Know Your Customer (KYC) process before their services can be used. For cryptocurrency exchanges, this is mandatory under anti-money laundering laws. As shown in Figure 4.6, the identity of the individual is provided to the service provider. The service provider transfers transactions created by the individual to the network. The blockchain network generally consists of network participants, which are described in more detail in Section 4.2.1. A public key found in the network can therefore be re-identified by adding the information of the service provider. As in Breyer vs the Federal Republic of Germany [Cou16b], it can be argued that the data on the blockchain is personal data for certain parties. In the general context of the GDPR, however, these data are sufficient to be classified as personal data because they are clearly attributable to an identifiable natural person.

Anderl and Schelling additionally argue that other protocol information, such as date and time, IP addresses, and wallet addresses, might be used to re-identify data subjects [And20, p. 100].

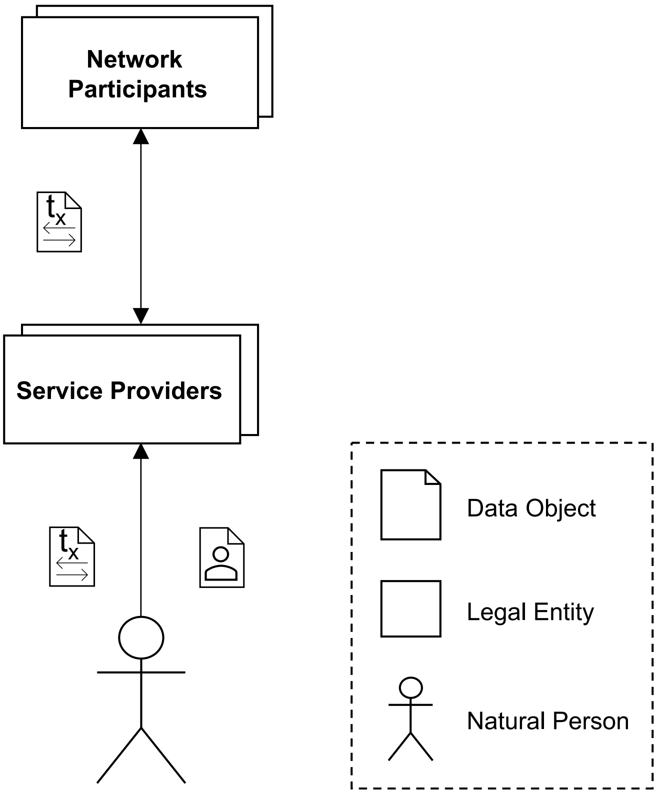


Figure 4.6: Identifiability through multiple data sources

Payload data depends on the application. It might contain personal data, mixed data or non-personal data. Section 3.5 described several common use cases on the blockchain. Several of them are storing personal data, according to the GDPR on the blockchain. Decentralized Identifier (DID) are generally personal data since their purpose is persona identification. Even though the technology exists for privacy enhancements, such as multiple identities, pseudonyms or tokens, DID are personal data by design. Registries, such as deeds and land registers, contain personal data. Automotive use cases such as vehicle registries, mileage and repair registries contain vehicle or usage information. [Boa20] lists data relating to 'driving style', 'wear and tear on vehicle parts' and Vehicle Identification Number (VIN) as indirectly identifiable data that can be related to a natural person and, therefore, personal data.

The conducted study and literature review draw a clear picture. It shows that personal data is processed on blockchains in protocols and payloads. This is a needed condition to constitute a processing operation according to GDPR Art. 4 (2) and, therefore, for application of the GDPR.

The risk limit in the sense of an 'acceptable risk' is only vaguely defined in the GDPR. Foremost controllers and processors must take the perspective of the data subject [Cou16a, Art. 35 (9)] and consider the “state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons” [Cou16a, Art. 32 (1)].

An interpretation of these paragraphs leads to the notion that a risk, assessed from the point of view of the data subject, can be accepted if the solution effort on the part of the controller and processor would be too high for the risk minimisation gained. Further restrictions and measures are prescribed should the residual risk be too high (see activity ‘Residual Risk Identification’)).

With an understanding of the context, the legal entities and the risk acceptance criteria, a further assessment can be done.

4.2.2 Risk Assessment

The initial risk identification searches for risks within the given context. The properties and issues with personal data have already been discussed in Section 2.1.3. In order to carry out a structured analysis of the risk, an identification of the processing activities of personal data on blockchains is necessary. On the one hand, this is necessary to check whether a risk exists and, on the other hand, to assess the extent of issues and the associated risks (*a_2*). Each risk is therefore identified (*a_2.1*), analysed (*a_2.2*) and evaluated against the risk acceptance criteria (*a_2.3*).

The risks are to be analysed taking the view of the data subject (see Section 4.1), therefore the rights of the data subject and the obligations of the data controller and processor will be taken as a structured basis for evaluating data protection risks regarding the data subject.

The principles of the GDPR, such as rights and obligations (see Section 2.3), cannot only be implemented by technology but must be carried out holistically in a company based on GDPR assessments. The bases for these are the entire company’s processes and data processing operations within these processes.

A single technology can make some principles more difficult or even make them impossible. Therefore, blockchains as technology must be examined in this regard.

The GDPR related risks are to be identified with the obligations and rights according to the template defined in Section 4.1. Evaluation based on transfer technology alone is difficult to achieve; nevertheless, the distinction between private and public blockchains is to be taken since the crucial properties of these are influencing the obligations and rights substantially. The results are therefore recorded in two separate risk registers (see Table 4.1 and Table 4.2).

A transfer in third countries [Cou16a, Art. 30 (1)] is hard to assess in the case of a public blockchain since the locations of nodes are unknown. The statistic referred to in Section 4.2.1 shows that in the case of the Bitcoin blockchain, nodes outside of the EU exist and that more than 50% of the nodes’ locations are unknown. In the case of public blockchains, it is very likely that a node outside the EU receives, and therefore processes [Cou16a, Art. 4 (2)], the data. Therefore the probability is high; the impact of transfer in a third country without guaranteed safeguards for data protection [Cou16a, Art. 46] is high since the rights of data subjects and measures for securing data cannot be guaranteed; the overall resulting risk priority is high (see *r_pu_1* on the risk register in Table 4.1).

Records of processing activities, including the aforementioned transfer to third countries, are needed to be fulfilled by the controller or joint controllers (cf. discussion regarding roles in the previous section) according to the transparency requirements defined by [Cou16a, Art. 30]. The problem with blockchain technologies' distributedness and decentralisedness are that the processing activities with data published on a public blockchain are unknown since anyone has access and can therefore process the data at will. Therefore records of processing activities [Cou16a, Art. 30 (1)] done on this data are hard to achieve for public blockchains in general. The impact on the freedom of the data subject is medium, but the probability is high. Therefore, the risk priority is rated as high (see *r_pu_2*).

On a private blockchain, the participants (nodes) are usually restricted by authentication and authorisation to participate in the blockchain network (e.g. process data). The access and, therefore, processing activities can be restricted and are usually known by the operator. The probability, though, is higher in such a distributed scenario than in simpler, non-distributed systems. Therefore, the risk in both transparency [Cou16a, Art. 30] is estimated as low (see *r_pr_1* and *r_pr_2* on the risk register in Table 4.2).

The technical and organisational obligations according to [Cou16a, Art. 32] such as data protection by design and data protection by default are different regarding private and public blockchains. Horizontal and vertical data minimisation in terms of quality and quantity (see Section 4.1.2) is difficult to achieve in both scenarios. To assess the risk occurring through data being shared extensively, the approach defined in Section 4.1.2 can be used.

The 'data amount' on the blockchain when sharing information is relatively high in relation to the actual data needed. Blockchain technology itself does not have functions for data requests and responses. The implementation of such functions can be done outside the chain on an additional layer or implemented in the main protocol stack of the blockchain itself (see Section 4.2.3).

Generally, the entire quantitative data set is stored on the blockchain, but all data fields in the qualitative sense are also stored on the blockchain. As explained in section 4.1.2, the risk is higher when more data in the quantitative and qualitative sense relating to a data subject is accessible to any party. Technical mechanisms that allow field-based data access can also be used to mitigate this risk (see the 4.2.3 section).

The relatability of data, especially transactions, is an issue in the blockchain technology field. Introducing artificial diversity through fake transactions and other noise will be discussed in the mitigation section. Also, data cannot be deleted or altered on blockchains, which leads to a possibly large amount of data about a data subject accessible at once on the blockchain. This might lead to the possibility of statistical reidentification (see Section 2.1.3) or narrow profiling in the case of pseudonymous data.

This directly relates to the third factor, namely the time span. The time data is persisted on a blockchain is usually infinite since it is one pillar of blockchain that data cannot be changed (immutability) once written. Also, time adds risk to certain mitigation strategies, such as encryption, since encryption might not be secure anymore in the future. Details will be discussed in the mitigation section (Section 4.2.3). Nevertheless, unintended or unauthorised disclosure or access [Cou16a, Art. 32 (2)] can occur when data has been secured, but the mechanism is being broken, such as the aforementioned reidentification of previously unidentifiable data or encryption. The risk of unintended disclosure or because of missing technical or organisational measures, in general, has a high impact on the data subject since too much information is disclosed to a third party. The probability of both risks is relatively high and therefore the risk priority is also high (see *r_pu_3* and *r_pu_4* on the risk register in Table 4.1). In case of private blockchains the probability would be the same but the impact lower (medium) since the possible receivers are limited (see *r_pr_3* and *r_pr_4* on the risk register in Table 4.2).

The *data protection impact assessment*, according to [Cou16a, Art. 35] is hard to achieve on a public blockchain because of the aforementioned missing knowledge of processing operations (*r_pu_2*). On a private blockchain, the data protection impact assessment can be easier, but processing activities are, because of the implementation of the communication channel, hard to achieve, though the risk of unknown processing operations is lower on private blockchains than on public blockchains. The missing data protection impact assessment directly has a low impact on the data subject but might have a medium to high impact indirectly because of missing information. Therefore the risk has been assessed with a medium impact for public and private blockchains and the probability has been assessed with medium in case of public and low with private blockchains (see *r_pu_2* and *r_pr_2* risk registers in Table 4.1 and 4.2).

A *data breach notification*, according to [Cou16a, Art. 34] is hard to achieve because of the identification of a data breach. On public blockchains, anyone can access the data and therefore try to break privacy-enhancing technologies such as encryption. The occurrence, or success, of such an attack will only be known when the damage has been done, i.e. the data has been published or used. A technical mechanism such as an intrusion detection system is not possible to implement on a public blockchain due to the absence of access data. A private blockchain might have the possibility to log data access to implement an intrusion detection system that analyses unusual activity that might indicate a data leak. In both cases, the awareness of publishing data to an external system and, in the case of a public blockchain, a publicly accessible system, can be assumed. Though, the missing of a data breach can have a high impact on the data subject since its information could be used for fraud. Therefore the probability in the case of a public and private blockchain has been assessed with low (see *r_pu_6* and *r_pr_6* risk registers in Table 4.1 and 4.2), but the impact is still rated as high.

[Cou16a, Art. 37] asks for the appointment of a data protection officer in certain cases (see Section 2.3.3). The appointment of a data protection officer itself does not change through the use of blockchain technology, but the management of contacts of controllers and processors is, in the case of a public blockchain, harder. In the case of a private blockchain register with access to the blockchain and its processing activities, a data protection officer should exist. These risks (see *r_pu_7* and *r_pr_7* on the risk registers in Table 4.1 and 4.2) are assessed with medium impact due to the indirect damage it could make and in both cases with low probability and resulting medium overall risk priority.

The *right of access* is a data subject's "right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access" [Cou16a, Art. 15 (1)] and further information specified in [Cou16a, Art. 15 (1)]. This right is easier to implement on a public than on a private blockchain. Every data subject has access to a public blockchain theoretically and might only need a program to access the data and to, interpret it, and export it to a "commonly used and machine-readable format" [Cou16a, Art. 15 (1)]. On a private blockchain, such export of data needs to be implemented in a separate program that has access to the private blockchain. In both cases, there can easily be an unintentional overuse of means to attack the identification and authorisation of the data subject. A Separate mechanism must exist that guarantees only authorised access. This risk has been assessed with medium impact since an active attack has to take place. The probability on a private and public blockchain has been assessed with medium (see *r_pu_8* and *r_pr_8*).

The *right to rectification*, according to [Cou16a, Art. 16] is hard to achieve on a blockchain (see *r_pu_9* and *r_pr_9*). This also applies to [Cou16a, Art. 17] the right to erasure or 'the right to be forgotten' (see *r_pu_10* and *r_pr_10*). The blockchain itself is, per design, immutable (see Section 3.3.1). This means that data that has been written once cannot be changed. Technical possibilities to mitigate this risk are discussed in the following section. Due to the very high probability, because of the basis of the technology and the high impact on many data subjects, both risks have been assessed with high impact and probability. Besides the problem of deletion of the data, another possible risk is the possibility of data recovery. Since the data has been published and accessed by unknown entities, it is impossible to prove whether the data has been deleted at all or deleted in a way that cannot be recovered (*r_pu_11* and *r_pr_11*). Both the probability and impact were rated medium. This is due to the fact that recovery is technically difficult and requires certain unlikely circumstances and that the impact would be confined to one data set rather than all data.

The *right to restriction* of processing according to [Cou16a, Art. 18] has similar issues as [Cou16a, Art. 16] and [Cou16a, Art. 17] though, an additional perspective needs to be added. Since the blockchain is a distributed system, a restriction of processing is harder to achieve, but the risk exists that a system still processes the information after a request for restriction has been made by a data subject. The impact for a data subject has been assessed with medium because the information is not accessed without consent, but the duration is merely extended. The probability on both public and private blockchains was assessed with medium (see *r_pu_12* and *r_pr_12* on the risk registers in Table 4.1 and 4.2) despite the fact that the likelihood that data will be further processed is inherently high since mere storage is sufficient as such. But, processing in the sense of further processing of the data can be obtained through self-limiting mechanisms in applications, as mentioned for backup mechanisms (see Section 2.3.5).

The *notification obligation*, according to [Cou16a, Art. 19] might be hard to achieve in the case of non-payload data meaning that the data does not belong to a certain application. In this case, the data subjects are unknown and, therefore, cannot be contacted. A public announcement as a replacement has the risk that the data subject does not receive the information, and therefore the impact would be medium (see *r_pu_13* and *r_pr_13*), and the probability is only given on a public blockchain with low, on a private blockchain very low.

The *right to data portability*, according to [Cou16a, Art. 20] is easier to achieve with a public blockchain, but the risk exists that the data format stored on the blockchain is unreadable for the data subject or the entity the data is intended to be transferred to. In the case of a private blockchain, a separate system would be needed to export the data in a certain format. Therefore the risk on a public and private blockchain has been assessed with high impact and low probability, resulting in a medium overall risk priority (*r_pu_14* and *r_pr_14*).

The *right to object*, according to [Cou16a, Art. 21], is again hard to implement since the blockchain is a distributed system, and therefore the communication about the objection to all processors is hard (see *r_pu_12* and *r_pr_12* on the risk registers in Table 4.1 and 4.2). Also, the immutability of the data could be an issue if the objection implies a deletion or anonymisation of the data (*r_pu_9* and *r_pr_9*). Both risks have already been assessed for other obligations (see risk registers in Table 4.1 and 4.2).

Automated decision-making and profiling according to [Cou16a, Art. 22] can occur on transaction information as well as payload. On a public blockchain, multiple entities store information on the same blockchain regarding possibly the same data subjects. This makes, combined with the openness of the blockchain, a medium risk of profiling. The probability of profiling on the public blockchain is high (see *r_pu_15*). A private blockchain would need a horizontal and vertical limitation mechanism to tackle this issue, but in general, the same issue exists on a private blockchain but with limited probability (see *r_pr_15* on the risk register in Table 4.2).

	GDPR Art.	#	Risk and Impact Description	Impact	Probability	Risk Priority
Obligations						
Records and transparency	Art. 30	r_pu_1	Unknown transfer in third country	high	high	high
		r_pu_2	Unknown processing activities	medium	high	medium
Technical and organisational obligations	Art. 32	r_pu_3	Data limitability (minimisation)	high	high	high
		r_pu_4	Unintended disclosure	high	high	high
Data protection impact assessment	Art. 35	r_pu_2	Unknown processing activities	medium	medium	medium
Data-breach notification	Art. 34	r_pu_6	Unidentified data breach	high	low	low
Appointment of a data protection officer	Art. 37	r_pu_7	Contactability of DPO	medium	low	low
Rights						
Right of access by the data subject	Art. 15	r_pu_8	Unintended access	medium	medium	medium
Right to rectification	Art. 16	r_pu_9	Possible immutability	high	high	high
Right to erasure	Art. 17	r_pu_10	Possible undeletable	high	high	high
		r_pu_11	Possibility of recovery	medium	medium	medium
Right to restriction of processing	Art. 18	r_pu_12	Slow dissemination	medium	high	medium
Notification obligation	Art. 19	r_pu_13	Contactability of subjects	medium	low	low

	GDPR Art.	**#**	**Risk and Impact Description**	**Impact**	**Probability**	**Risk Priority**
Right to data portability	Art. 20	r_pu_14	Readability of format	high	low	low
Right to object	Art. 21	r_pu_12	Slow dissemination	medium	medium	medium
		r_pu_9	Possible immutability	high	high	high
Automated decision-making & profiling	Art. 22	r_pu_15	Unknown data sources	medium	low	low
		r_pu_8	Unintended access	medium	medium	medium

Table 4.1: Risk register of the assessment of public blockchains (r_pu)

	GDPR Art.	#	Risk and Impact Description	Impact	Probability	Risk Priority
Obligations						
Records and transparency	Art. 30	r_pr_1	Unknown transfer in third country	high	low	low
		r_pr_2	Unknown processing activities	medium	low	low
Technical and organisational obligations	Art. 32	r_pr_3	Data limitability (minimisation)	medium	medium	medium
		r_pr_4	Unintended disclosure	medium	high	medium
Data protection impact assessment	Art. 35	r_pr_2	Unknown processing activities	medium	low	low
Data-breach notification	Art. 34	r_pr_6	Unidentified data breach	high	low	low
Appointment of a data protection officer	Art. 37	r_pr_7	Contactability of DPO	medium	low	low
Rights						
Right of access by the data subject	Art. 15	r_pr_8	Unintended access	medium	medium	medium
Right to rectification	Art. 16	r_pr_9	Possible immutability	high	high	high
Right to erasure	Art. 17	r_pr_10	Possible undeletable	high	high	high
		r_pr_11	Possibility of recovery	medium	medium	medium
Right to restriction of processing	Art. 18	r_pr_12	Slow dissemination	medium	medium	medium
Notification obligation	Art. 19	r_pr_13	Contactability of subjects	medium	low	low

	GDPR Art.	**#**	**Risk and Impact Description**	**Impact**	**Probability**	**Risk Priority**
Right to data portability	Art. 20	r_pr_14	Readability of format	high	low	low
Right to object	Art. 21	r_pr_12	Slow dissemination	medium	medium	medium
		r_pr_9	Possible immutability	high	high	high
Automated decision-making & profiling	Art. 22	r_pr_15	Unknown data sources	medium	low	low
		r_pr_8	Unintended access	medium	medium	medium

Table 4.2: Risk register of the assessment of private blockchains (r_pr)

The principle of transparency in front of the data subject and the data protection authority (see Section 2.3.3) can be supported by public blockchains. The principle of immutability and transparency allows anyone to assess the data and logic (contracts) on the blockchain. This does by far not mean that all transparency obligations are automatically fulfilled, but users can check on automated decisions and (categories) of data being processed on-chain. When looking at Table 2.1, three of these transparency requirements, in contrast, are not easily fulfilled: categories of recipients, transfer to third countries, and storage period. The recipients of data published on a public blockchain are not known. One could argue that publishing data on a blockchain makes them publicly available, and therefore, special rules apply. The transfer in third countries cannot be answered for sure, whereas one could argue that all known public blockchains have a node outside the European Union. The storage period is the reversal of the obligation to delete data after the purpose has expired. Since data cannot be deleted from a blockchain after it has been written (immutability), it cannot be deleted at all. In the course of the obligations of the processor or controller, however, it can be stated that the storage period can only be determined for the storage of the processor or controller itself due to the publication of the data on the blockchain. Section 4.2.3 discusses possible solutions to these risks.

Technical and organisational measures [Cou16a, Art. 32] regarding the data processing activities that need to be implemented. The CIA Triad properties (see Section 2.3.3 for description), namely confidentiality, integrity and availability, are partially hard to implement. Confidentiality, i.e. preventing disclosure to unauthorised parties, requires special solutions for public blockchains, as the data cannot be changed, and therefore security measures such as encryption cannot be subsequently improved; data once published with encryption that is later broken cannot be retracted and is therefore exposed. The integrity of data, in contrast, is automatically given by the blockchain property of immutability. Data can (almost) not be changed by third parties. Availability, i.e. the accessibility of data, is guaranteed by the distributed nature of blockchain technology. Denial of service attacks are hard to perform against a full blockchain. All these technical measures must also be applied to the applications that use blockchain technology and interact with the customer (e.g., a portal), and organisational measures adapted to them must be taken.

Implementing privacy by design is also hard. Privacy by design includes data minimisation (see Section 2.3). Blockchain as a data transfer system does not allow for horizontal or qualitative nor vertical or quantitative (see 2.3.6) restricting of data. Every system communicating over the same blockchain receives the same information (without any further extension).

With regard to the rights of data subjects, there has been the most discussion in the scientific community. On the one hand, public blockchains allow easy access to stored data and, therefore, might directly fulfil the right to data portability [Cou16a, Art. 20]. Article 20 states that the data subject has the right to receive his or her data in a machine-readable format, which blockchains are in general since the data structure is defined (see Section 3.3).

The right to erasure of personal data is technically impossible to implement since data once written on a blockchain cannot be deleted. It is discussed whether it could be sufficient to delete information in the sense that the identifiers to the published information are deleted. Also, other solutions to the issue have been discussed, including encryption and deletion of the key. These solutions will be discussed in the following section.

4.2.3 Risk Treatment

The stated problems seem to be technical or organisational solvable or avoidable. A literature review resulted in a selection of solutions that are evaluated and discussed in the following sections. These are then applied to the risks identified in the previous sections.

Redactable Blockchain

The most straightforward idea to tackle GDPR related issues is to make blockchains editable. Editable blockchains are an extension of the original blockchain concept that allows for modifications of the blockchain's contents by authorised participants. These changes can be made unilaterally or in response to an agreed-upon consensus (see Section 3.4).

[LCNJ20] designed a 'GDPR Compliant' Blockchain that allows for the deletion of data with the help of Chameleon Hashes and Attribute-Based Encryption (ABE). Chameleon hash functions, also called trapdoor hash functions, are hash functions that have a trapdoor. This trapdoor allows the hash creator to create a collision, i.e. two hashes (digests) with a different input but the same output [CDK+17]. However, chameleon hash functions are collision-resistant as long as the secret key belonging to the trapdoor is not known. This can be used to allow the creator of a block to change the block later.

[Kup20] describe a concept allowing multiple 'contexts' to be created on one blockchain, whereas the blockchain is less a list but a tree with a branch for each context. This allows for the deletion of a context without influencing another context but reduces security (weight). [LYOK19] have a similar idea but use truncated hash values and sidechains to achieve the same.

While there are many potential applications for editable blockchains, there are also some potential risks associated with them. One concern is that unauthorised changes could be made to the blockchain, potentially compromising its integrity. Additionally, it is important that any changes made to the blockchain be consistent with its original purpose and design principles. This feature would allow for more flexible governance structures but would remove the property of an immutable blockchain. The consequences of such a change would be far-reaching and open possibilities of censorship and fraud.

The concept of a redactable blockchain can help tackle the 'right to rectification', 'right to erasure', 'right to object' and specifically the risks 'possible immutability' (r_9), 'possible undeletable' (*r_10*). It, at the same time, adds new risks that will be discussed in the residual risks section.

Differential Privacy and Noise

Differential privacy attempts to develop mechanisms that prevent individuals from being identified in a dataset. Specifically, Le Ny [LN20] formulates that a dataset regarding one person added to a known dataset does not provide any new information about that person, but as much information exists about them as if they were not in the dataset. Most differential privacy approaches try to add data that does not interfere with the requested query. With blockchains, there is usually no request-response system; all data is accessible to everyone. Therefore noise can only be added to the full data set - still with the limitation that it does not interfere with the computation done.

Noise is an unwanted, unexplained, random variability in data. Too much noise in data makes the data unreliable. Adding statistical noise to data sets can make the identification of individuals more complicated whilst not influencing the results of computation done on the data.

Adding noise can be anything from random data to specific patterns that are unlikely to occur naturally. One of the most common types of noise is generated with a Pseudorandom Number Generator (PRNG). This type of noise is created by mathematical algorithms and is very difficult for humans or computers to predict.

Monero tries to obfuscate the transactions by inserting multiple signatures in one transaction with the help of ring signatures [VS13]. Only one of the signatures is the signature needed for the transaction; the others are randomly generated to reduce the chance of identifying the path of transactions.

Differential privacy by adding noise is promising because the Article 29 Working Party has already recognised that after the removal of apparent personal data and quasi-identifiers, the addition of noise may be an acceptable form of anonymisation in their opinion [Par14].

Waste tracking applications on the blockchain [SPS+19], for example, can add noise to location data without compromising the goal of avoiding the loss of waste and being able to track it down in the event of an incident. Genomic data can have noise added to it after encryption to make it securely available to researchers [PKS21].

The concept of differential privacy and noise can help tackle many risks since a higher degree of diversity in data can help mitigate certain risks in terms of reduction of the probability or impact. If the data can be made completely anonymous, it would be not only a mitigation but a full avoidance. This especially (but not only) applies to the following risks: 'unknown transfer in third country' (*r_1*) would not apply in the case of anonymous data; 'unknown processing activities' (*r_2*) would not apply since it does not state a processing activity on anonymous data; 'unintended disclosure' (*r_4*) would be mitigated (from the perspective of the data subject) since the disclosed data would not be connectable to a data subject; 'unintended access' (*r_8*) has the same reasoning.

Also, the ‘right to rectification’, ‘right to erasure’, ‘right to object’ and specifically the risks ‘possible immutability’ (*r_9*), ‘possible undeletable’ (*r_10*) would be avoided since the rights of a data subject are limited to personal data.

State channels

A state channel is a way of making transactions between two parties more efficient. Normally, when two parties want to make a transaction, they would have to broadcast their transaction to the entire network, which would eventually record the transaction on the blockchain. State channels, as a technology, allow for so-called ‘off-chain’ transactions between two or more participants. These transactions are not broadcast to the entire network (‘on-chain’) but rather are handled privately between the participants involved. This makes the process faster and more private. They are currently being used for the Bitcoin blockchain (i.e. lightning network).

Before being able to send a transaction over a payment channel, a transaction is funded as security. This prevents arbitrarily quitting the payment channel to one party’s advantage. This opening of a payment channel is done as an on-chain transaction that needs a transaction fee. After a payment channel has been opened between two parties, the transactions made in the payment channel are off-chain, meaning that they are not recorded on the blockchain. Therefore no transaction fee has to be paid in general, and then the transaction is instantaneous and private. When the channel is closed, the final state of the channel is recorded on the blockchain, again charging transaction fees. In case of a dispute, both parties can close the payment channel receiving their last committed value. The commitment is reached by constantly exchanging time-locked transactions and revocation keys for previous transactions.

One of the main advantages of state channels is that they can greatly reduce blockchain congestion. When too many people are trying to make on-chain transactions, the network becomes congested, and processing times increase. State channels can help to alleviate this congestion by allowing for private transactions. This also helps to reduce fees since there is less competition for space on the blockchain (see Section 3.4).

Another advantage of state channels is that they allow for more privacy than on-chain transactions. On-chain transactions, such as the opening and closing transactions, are publicly recorded on the blockchain and can be viewed by anyone who chooses to do so. Transactions within the channel, however, are kept private between the participants involved. Only in case of an exception or when closing the channel an on-chain transaction is published.

While state channels offer many advantages, there are some potential drawbacks as well. One potential drawback is that state channels can be complex and difficult to set up correctly. If not implemented properly, they can lead to errors and financial losses for those involved in the transaction process.

The concept of state channels can help tackle risks related to the property of a blockchain in that data is published to many partially unknown recipients. The ‘Unknown processing activities' (*r_2*) and 'Unknown transfer in third country' (*r_1*) are an example of two risks that can qualify the unknown and therefore avoid the risk. It directly tackles the risk of 'data limitability (minimisation)' (*r_3*) by limiting the recipients. The limitation of recipients allows for quantitative and qualitative limitations since a node can choose what data to send to a certain other node. 'Unintended disclosure' (*r_4*) cannot be fully avoided, but a certain risk can be mitigated. Also, access activities can be qualified, resulting in a reduction of the risk 'unintended access' (*r_8*). This qualification of access can also result in faster dissemination of restrictions and, therefore, a reduction in the risk 'slow dissemination' (*r_12*) and 'unknown data sources' (*r_15*).

Ring signatures

Ring signatures are a cryptographic tool that allows for transactions to be signed by multiple parties, such that the actual signer is hidden. Ring signatures were originally proposed by Ron Rivest, Adi Shamir, and Yael Tauman in 2001 [RST01]. They are used in the digital currency Monero as well as other applications.

A ring signature is created by combining the public key of a sender with a group of other public keys, or 'siblings'. The siblings can be from past transactions, or they can be generated randomly. When a ring signature is created, it is impossible to determine which of the siblings was responsible for signing the transaction. This provides privacy for both the sender and receiver of a transaction.

One potential use for ring signatures is in voting systems [KKP^{+}19]. In a traditional voting system, each voter's ballot is tied to their identity and can be traced back to them. With ring signatures, voters could cast multiple ballots without anyone being able to trace them back to their original vote. This would help protect voter privacy and prevent vote-rigging.

The concept of ring signatures can help tackle risks related to the relatability of pseudonyms. Signatures are, by design, relatable to a person and are, therefore, pseudonymous. Ring signatures help to obfuscate the relation by probability. They can help reduce the risk of 'unknown transfer in third country' (*r_1*) with less relatability to a data subject. Also, 'unintended disclosure' (*r_4*) and 'unintended access' (*r_8*) would be reduced since the disclosed data would be less relatable to data subjects.

Onion Routing

Onion Routing is a general purpose infrastructure for private communication over a public network [SGR99]. It is implemented utilising a low-latency mix network [DMS05]. Data streams are routed through a series of onion routers, each of which knows only the identity of the next router in the chain but not the ultimate destination. Data by the sender is encrypted in multiple layers, each layer being removed in every step of the transfer process[GRS96]. The vegetable onion is used as a metaphor since each router represents a different layer of an onion. The first router knows only the outermost layer of the message, which is then forwarded to the next router. This router then again removes the outer layer and forwards the message to the next router, and so on. This happens until the final recipient removes the last layer and reads the message. If a sender wants to communicate with a service outside the Tor network, an exit node is used as a gateway to the open Internet.

Onion routing was initially developed by Michael G. Reed and Paul F. Syverson from Naval Research Laboratory for use in protecting U.S. intelligence communications online [RSG98] [Smi13]. Onion routing was further developed into Tor (The Onion Router) software by Roger Dingledine and Nick Mathewson from The Tor Project [4].

The purpose of Tor was to protect government communications, but now it is used worldwide by people who want to keep their Internet activities private from websites they visit, companies they do business with, or governments that want to track them. Tor also enables users to access the 'deep web' and 'dark web', which is a collection of websites that are not indexed by standard search engines and accessible with 'onion' domains, whereas the dark web is a small subset of the deep web which is intentionally hidden.

This layered encryption makes it difficult for anyone who may be monitoring traffic between two points to determine where a message originated or where it is destined. It also prevents anyone from knowing how many layers are in use or even how many routers are in communication with each other.

Anyway, there have been several attacks against Tor [EHK$^+$16]. Statistical attacks, such as end-to-end traffic correlation attacks try to de-anonymise the sender by correlating the sent traffic (size and other metadata) to the traffic received at a service [Dan04, LRWW04, ZFG$^+$04, SS03]. Congestion attacks try to identify the routers between the sender and the service by sending large amounts of data to routers and watching for the data stream between the sender and the service to get slower [EDG09].

Onion routing is a promising technology for reducing the information about the sender of messages (i.e. the identity). It has been practically used on hypertext and blockchain communication. This technology, on the one hand, reduces information about a data subject on a node and, on the other hand, tries to enlarge the diversity in the data. The information being hidden is only meta-information or routing information, not payload information. Whether the risk of identification can be fully avoided or it only represents mitigation is discussable.

The concept of onion routing can help tackle risks related to unwanted disclosure about a node itself. It is a PET to minimise data and reduces the risks 'data limitability (minimisation)' (*r_3*) and 'unintended disclosure' (*r_4*).

Controversially the concept also makes the risk of 'Unknown transfer in third country' (*r_1*) worse since the country of a node is actively hidden. This also relates to all risks where direct contact with a node is needed: 'contactability of DPO' (*r_7*) and 'contactability of subjects' (*r_13*).

[4] https://www.torproject.org

Pruning

A way to reduce data stored on a node is pruning. Pruning is the process of removing unnecessary or redundant data from a system in order to improve performance and free up storage space. Pruning is especially important for large systems, such as databases or file systems, where there is a lot of data to manage. By removing unnecessary data, pruning makes it easier to find the information that is needed and reduces the risk of data leaks.

Pruning helps data protection by making sure that only the needed information (minimisation) is stored on the system. It also makes the system more efficient so that data can be accessed and updated more quickly.

On the Bitcoin blockchain, pruning options are limited since most data stored by a full node is needed to verify new blocks. The unspent transactions, which are most important on the underhand, are usually stored in a structured, local database to allow for fast access.

Pruning in Ethereum is accomplished by deleting data from the local storage. When a node receives a new block, it verifies that block against its local copy of the blockchain. If there is no conflict between the two versions, then the node can delete its old copy of the block from local storage (See Section 3.4 on uncle blocks). This process is repeated for all subsequent blocks until only the most recent block is retained on the local storage.

Pruning improves performance by eliminating redundant data processing on nodes. It also helps to prevent nodes from becoming filled with data and slowing down. By removing older blocks from circulation, pruning also helps to reduce contention for scarce resources like disk space and bandwidth and, foremost, to reduce risk by storing unneeded personal data.

The concept of pruning can help to reduce the risks associated with the mass storage of data that may no longer be needed. It minimises data and reduces the risks 'Data limitability (minimisation)' (*r_3*) and 'unintended disclosure' (*r_4*). It also indirectly reduces the risks associated with data breaches 'unidentified data breach' (*r_6*) and 'unintended access' (*r_8*). Also, the possibility of recovering information regarding a data subject is lowered (compared to other techniques) 'possibility of recovery' (*r_11*).

Homomorphic encryption

Homomorphic Encryption (HE) is a technique for performing arithmetic or logical operations on encrypted data without decrypting the data first [Lau21]. This allows a third party to compute data without knowing the content of the data - hence the data is protected.

Homomorphic encryption has existed for a long time, but the first Fully Homomorphic Encryption (FHE) was proposed by Craig Gentry in 2009 [Gen09]. There are several categories of HE: partial, levelled fully and fully. Whereas the first two allow limited operations (by size or nature of the operation) latter allows arbitrary computations on the data [Lau21]. FHE allows to perform arbitrary operations on encrypted data, including addition, subtraction, multiplication, and division.

There are a number of applications for privacy-preserving computations, including biomedical research, financial analysis, and government intelligence gathering. In each of these cases, it is important to protect the privacy of the data being processed. Homomorphic encryption provides a way to do this while still allowing meaningful computations to be performed.

There are some practical challenges involved in implementing homomorphic encryption schemes securely. Some of these have been overcome in recent years with advances in cryptanalysis and efficient implementation techniques, while others are still being worked on. One challenge is the size of the HE's ciphertexts which is 10 times to 10^4 times higher than in standard schemes such as AES [AOSV22]. Also, the runtime of HE schemes must be improved [AE16]. As a result of the progress, there is increasing interest in applying homomorphic encryption techniques across a range of applications, including blockchain.

Homomorphic encryption is also important for blockchain because it enables the execution of smart contracts on encrypted data [SA21]. This means that the data can be processed without revealing any information about the underlying contract to the party performing the computation. This is important for privacy-preserving smart contracts because it allows the contract to be executed without revealing any information about the contract to the party performing the computation.

Homomorphic encryption has been used in blockchains to hide the transaction amounts [WSL+19]. Monero uses homomorphic encryption in their range proof system based on Bulletproof replacing the larger ring signatures [Pro] (see Section 4.2.3 for details on zero-knowledge proofs). Zerocoin[5], an improved version of Bitcoin, uses homomorphic encryption to hide coins [WSL+19].

Zilliqa[6] plans to use homomorphic encryption to create privacy-enhanced Decentralized Applications (DApps); meaning that data can be stored encrypted on their blockchain, and a smart contract can compute on that data without encrypting it. Currently, they are using Schnorr signatures on Elliptic Curves [WSL+19]. Schnorr signatures are named after their inventor, Claus-Peter Schnorr. They are a type of digital signature that is based on Elliptic Curve Cryptography (ECC). Compared to the previously used standard ECSDA signatures, Schnorr signatures on the same curve (secp256k1) offer several advantages, including provable security, non-malleability, and linearity [ELOP20]. In multi-signature scenarios, they are also shorter and therefore use less storage than naive concatenation signatures [MOR01]. This makes them ideal for use in cryptocurrency transactions.

Overall, homomorphic encryption represents an important advance in cryptography that has the potential to revolutionise blockchains.

[5] https://zerocoin.org

[6] https://www.zilliqa.com/

Homomorphic encryption can hide information from third parties, even if they need to compute that data. The applications of homomorphic encryption are currently limited but are expected to increase. They can contribute to a general reduction of data risk (‘mitigation’) because data is not only encrypted during transfer or storage but also during processing. As long as there is a possibility of decryption, this data is still personal in any case. A further discussion on encrypted data takes place in Section 3.

It allows the data to be less relatable (cf. diversity), which means that it is harder to link the data to an entity, but it does not make it anonymous in the sense of the GDPR.

The concept of homomorphic encryption can help to reduce the data being transferred (including private data). It potentially allows to avoid the transfer of personal data completely. ‘unknown transfer in third country’ (*r_1*) would not apply in case only anonymous data is transferred. ‘unknown processing activities’ (*r_2*) would not apply since it does not state a processing activity on anonymous data. It therefore avoids the risks ‘Data limitability (minimisation)’ (*r_3*) and ‘unintended disclosure’ (*r_4*). It also reduces the risks associated with data breaches ‘unidentified data breach’ (*r_6*) and ‘unintended access’ (*r_8*). Also the ‘right to rectification’, ‘right to erasure’, ‘right to object’ and specifically the risks ‘possible immutability’ (*r_9*), ‘possible undeletable’ (*r_10*) would be avoided.

Hash Functions

Hash functions take inputs of arbitrary size and fit them in a fixed-size data structure (digest) [Sie]. The digest size depends on the specific hash function used. Even though most hash functions produce a fixed-size output, some functions, such as ‘sponge functions’, can produce a given output length. A sponge function takes “a variable-length input and produces an infinite-length output” [BDPVA07]. The National Institute of Standards and Technology (NIST) Secure Hash Algorithm-3 (SHA-3), which is based on the KECCAK algorithm uses such a sponge function [D$^+$15]. As a one-way function, the hash function is easy (cp. computational complexity) to compute but hard to reverse. Unkeyed hashes map data to a fixed-size digest and are easily computable. A cryptographic hash function must have all properties of an unkeyed hash function and, additionally, pre-image resistance, second pre-image resistance, and collision resistance [MvOV18, p. 323]. Whereas *Pre-image resistance* means searching for a message with the same hash of an existing hash must be difficult ($h = hash(m)$). *Second pre-image resistance* means that searching for a message with the same hash as an existing message must be difficult ($hash(m) = hash(m2)$). *Collision resistance* means searching for any two messages with the same hash must be difficult ($hash(m) = hash(m2)$).

Cryptographical hashing does not automatically make secure pseudonymous data, especially not anonymous data. Reidentification of data might be possible by several techniques, such as rainbow tables on low entropy datasets or linking information [Eur19a]. These can be mitigated by other techniques, such as key stretching (salting) and entropy monitoring.

The hashed (and eventually salted) data is still connected to the data, even if only unidirectionally (see Figure 4.7). The 'Email' field in this figure is bidirectionally connected to the input data since it has not been changed. A data entry with this email address would be directly connected to the author. On the contrary, the hash (digest) of the email is, first of all, not directly usable and unreadable, but it is, foremost, not reversible. This means that anyone who knows that the hash has been produced with the standardised SHA1 algorithm would be able to create the same hash from the given email address, but no one (with the exception mentioned above because of attacks) can compute the email when only having the hash. The ethics guidelines of the European Commission for H2020 fundings summarise the same view by stating that "pseudonymisation entails substituting personally identifiable information (such as an individual's name) with a unique identifier that is not connected to their real-world identity, using techniques such as coding or hashing. However, if it is possible to re-identify the individual data subjects by reversing the pseudonymisation process, data protection obligations still apply. They cease to apply only when the data are fully and irreversibly anonymised" [Com18]. Only anonymous data that cannot be re-identified classify for an exemption from the GDPR. The last column of Figure 4.7 shows an anonymisation technique according to the GDPR. Factually, it is a de-identification technique (see Section 2.1.3). This information does "not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable"[Cou16a, Rec. 26] and is therefore not regulated by the GDPR.

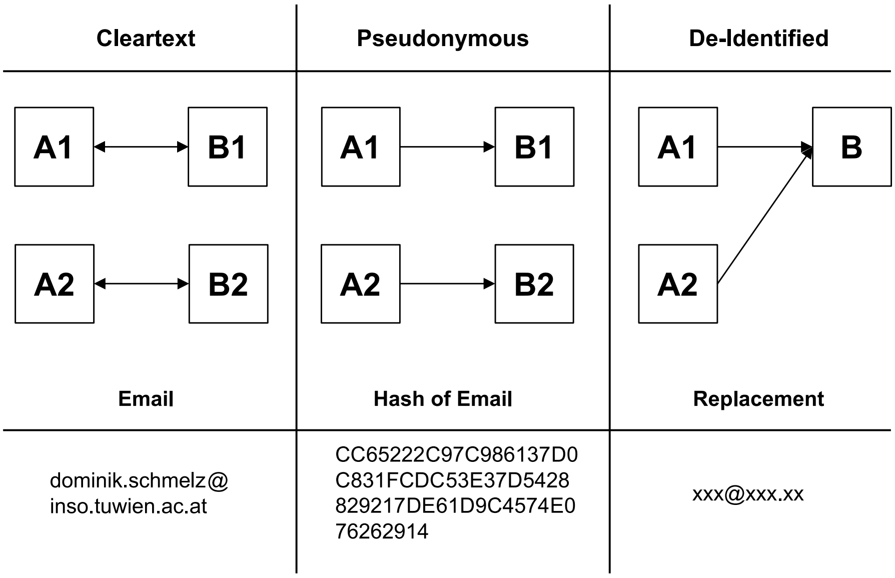


Figure 4.7: Pseudonymous Data vs De-Identified Data

Hash functions can, therefore, only be seen as a 'mitigation' of the risk entailed with non-pseudonymous data, but not an 'Avoidance'. It reduces the relatability (cf. diversity) of the data, which means that it is more difficult to link data to an entity, but it does not make it anonymous in the sense of the GDPR.

The concept of hashing can help to reduce the relatability and therefore reduce risks associated with disclosed data. They can help reduce the risk of ‘unknown transfer in third country’ (*r_1*) with less relatability to a data subject. Also, ‘unintended disclosure’ (*r_4*) and ‘unintended access’ (*r_8*) would be reduced since the disclosed data would be less relatable to data subjects.

Stealth Addresses

Stealth addresses are a concept of wallets (see Section 3.3.4) to allow for a receiving party to generate as many addresses as needed to disguise the association of multiple transactions to one recipient. Usually, a wallet creates one address per transaction. As mentioned in Section 3.3.4 Child Key Derivation (CKD) allows to generate multiple keys from one seed. BIP32 [Wui12] defines a function that generates three keys from one seed. Recursively called a tree of keys can be generated, forming a Hierarchical Deterministic (HD) wallet. This has a privacy-enhancing effect. A third party watching the public transactions would only see several transactions, not being able to identify that they are actually to the same recipient. This may seem quite beneficial in the short term, but in the long run, a third party may see merging transactions (see Section 3.4.4) that could eventually identify the receiving party (see Figure 4.8). A value emitting party broadcasts a transaction ($tx_{1.1}$) with a funding ($tx_{1.1}$ input0) and a receiving address ($tx_{1.1}$ output0)[7]. Later the same receiving party should receive a value. Instead of using the same receiving address (which technically would be possible), the receiving party generates a new address ($tx_{1.2}$ output0) to disguise that the ‘output0’ of each transaction belongs to the same receiving party. The issue with this technique arises when the receiving party wants to spend the received values ($tx_{2.1}$ and $tx_{2.2}$). This is due to the fact that a wallet will eventually use both received values to fund transactions that can be related to each other. In a simple scenario, it, for example, receives two times 0.3 BTC and needs to fund a 0.5 BTC transaction. A simple analysis of tx_3 shows that both transactions probably belong to the same entity. The de-identification attempt failed. An advanced wallet will try to mitigate this by not using the funding of the same emitting party. This will lower the probability of reidentification.

This technique is, therefore, to be classified as a ‘mitigation’ of the risk entailed since there is only a shift in the probability of identification. It does not make the data anonymous and is therefore not an ‘Avoidance’. It reduces the relatability (cf. diversity) of the data, making it more difficult to link transactional data (behaviour) to an entity.

The concept of stealth addresses can help to reduce the relatability of a behaviour (e.g. transactions) to a data subject. It, therefore, reduces risks associated with disclosed data. It is a PET to minimise data and reduces the risks ‘data limitability (minimisation)’

[7] Note: For ease of explanation the wording ‘address’ is used. For details on actual transaction scripts refer to Chapter 3

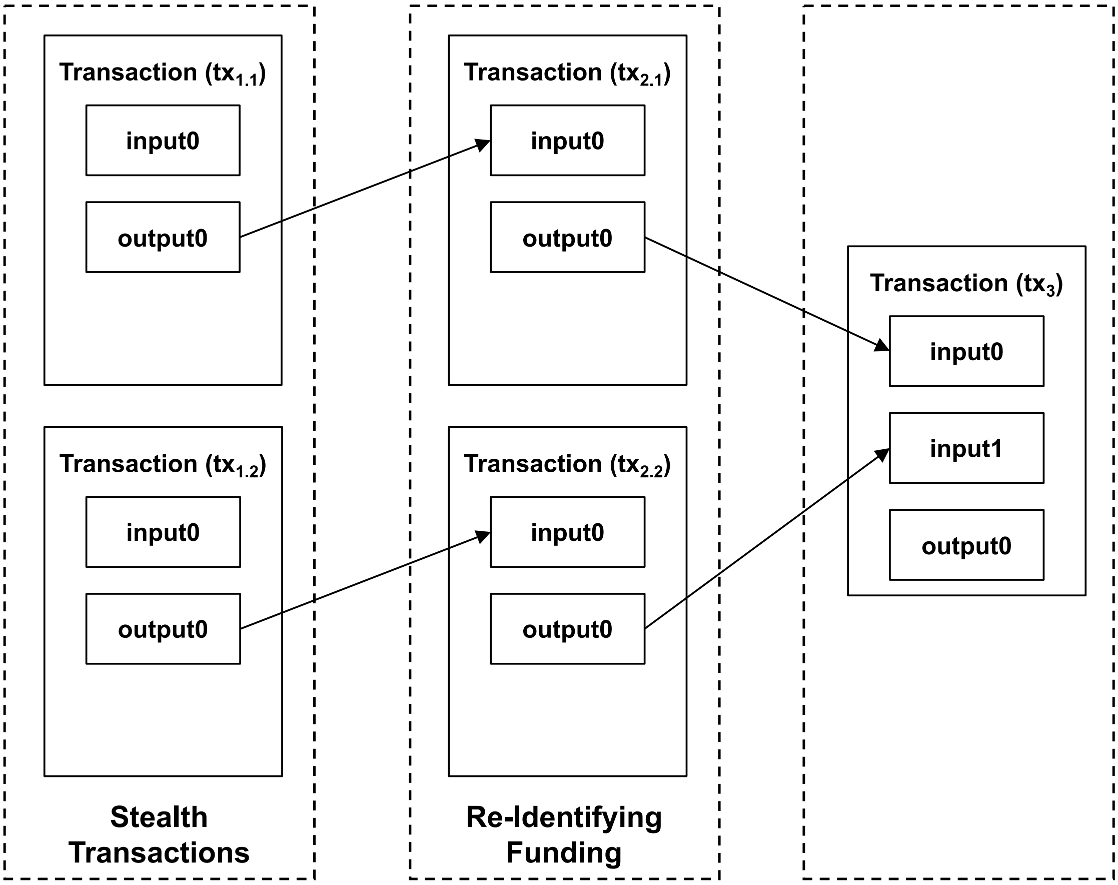


Figure 4.8: Stealth Addresses

(*r_3*), 'unintended disclosure' (*r_4*) and 'unintended access' (*r_8*) would be reduced since the disclosed data would be less relatable to data subjects.

It does not influence the 'contactability of subjects' (*r_13*) since only multiple identities are assumed instead of anonymising the identity.

Off-Chain Identifiers

The idea of off-chain data is to store direct personal data (see Section 2.1.3) in a separate database and only link the entries to the on-chain data with an identifier. This process can be classified as pseudonymisation according to the GDPR. The GDPR defines pseudonymous data as data that can "no longer be attributed to a specific data subject without the use of additional information" [Cou16a, Art. 4 (5)] and adds the requirement that the additional data needs to be kept separately. Otherwise, the data might be classified as indirect identification data. Pseudonymisation is a valid privacy technique applied to personal data but does not make data worth protecting since the link to the data subject is still very intact, and therefore a risk for the data subject does still exist, even though the risk has been lowered. The risk, however, is that reidentification can occur by statistics or legal or illegal access to the privacy guarding information. Therefore the GDPR recognises pseudonymisation as a 'safeguard' [Cou16a, Art. 6 (4)], but still classifies pseudonymous data as 'personal data'. Storing personal data off-chain and linking it with an identifier such as a hash or a foreign key linking to an artificial primary key or a simple random identifier helps, therefore, to protect data, but does not solve the issue on hand, risking the privacy of data subjects.

Section 2.1.3 explained that the field of medical trials and related legislation such as the Health Insurance Portability and Accountability Act (HIPAA) had the same problems before blockchains existed. Proper protection against reidentification is needed before publishing medical data. Differential privacy mechanisms that add noise depending on the size of the resultset, as suggested by [DMNS06], are not feasible when data is published on a blockchain. This data must be checked for diversity and identifiability and accordingly aggregated, or noise is added.

The concept of off-chain identifiers can help to limit the group of people entitled to depseudonymisation. Relatability is globally unrestricted, which means that an entity that has access to both information (off- and on-chain) has no restrictions. Locally, however, the availability is limited. An entity that only has access to the on-chain data cannot relate it to a person. Should this be sufficient for the use case, it would minimise the following risks: reduction of the risk 'unknown transfer in third country' (r_1) with less relatability to a data subject. Also, 'unintended disclosure' (r_4) and 'unintended access' (r_8) would be reduced since the disclosed data would not be relatable to data subjects. The concept does not anonymise the data, so deleting the off-chain data is not sufficient but would reduce the risk. Therefore the 'right to rectification', 'right to erasure', 'right to object' would be enhanced and specifically the risks 'possible immutability' (r_9), 'possible undeletable' (r_10) would be reduced.

Encryption Instead of Deletion

The idea of encryption instead of deletion describes a solution to the issue that personal data on the blockchain cannot be deleted once written. Seeing blockchain as a communication channel makes sense since only those who know the key to decrypt the data can read the data. In the context of deletion, one problem occurs: the communication channel is persistent. Once the key has been deleted, no one can seemingly access the data. Since the data is persisted, someone in the future could and probably can access the data. When looking at the risk, usually the worst case for encryption is cited, such as quantum computing or a bug in encryption schemes. The question whether the encrypted data is safe is easier to answer when looking at the current key-length recommendations by NIST [Bar20a] and BSI [BSI21]. The theory behind encryption is that the encryption process is easy to calculate, and the breaking of the encryption is too time-consuming with current computing power, but not impossible. Usually, the time needed for breaking a key by guessing grows exponentially by key length. Naturally, the key length must grow when computers get more powerful. Therefore the recommendations regarding key length, such as NIST [Bar20a] and BSI [BSI21] try to predict the advances and update the recommendations accordingly.

The selection of a key length that is suitable to protect private data that will stay on a blockchain forever is, therefore, impossible. Encrypted data stored on the blockchain will get cracked eventually. According to the GDPR, encryption is a 'safeguard' [Cou16a, Art. 6 (4)], but encrypted data still 'personal data'. Therefore the idea of deleting data by deleting the key is not feasible.

The concept of encryption (as a substitute for deletion) cannot fully solve the objection, restriction, and deletion issues but can help to reduce the risks associated with personal data. Therefore it is merely protecting 'unintended disclosure' (*r_4*) and 'unintended access' (*r_8*) since the disclosed data would not be immediately relatable to data subjects. The concept does not turn the data anonymous. Therefore, deletion of the keys would not be sufficient but reduce the risk. Therefore the 'right to rectification', 'right to erasure', 'right to object' would be enhanced, and specifically, the risks 'possible immutability' (*r_9*), 'possible undeletable' (*r_10*) would be reduced. Though the use of encryption instead of deletion does higher the risk of recovery of deleted data (*r_11*).

Bloom Filters

Within the blockchain, network nodes share the blockchain state (i.e. transactions) over a peer-to-peer network (see Section 3.4.1). Nodes that want to receive transactions from other nodes can use filters to only receive transactions they are interested in. These filters, on the one hand, limit the bandwidth usage of a node but, on the other hand, reveal information about them (i.e. what transactions they are interested in). Bloom filters are designed to add probability to the filter so that a node receives transactions it wants to receive (positive), but also some it did not want to receive (false positives) without filtering transactions the node wanted to receive (false negatives) [Mit09]. These false positives usually can be configured so that nodes wanting more privacy could reduce the loss of privacy by setting a higher false-positive rate. [GCKG14] showed that the rate must be quite high to ensure plausible deniability and that simple attacks, such as multiple connections to a node, can limit the efficiency of the filter, making bloom filters irrelevant.

From a data protection view, these filters reduce the data amount stored on the one hand and add diversity to data others have about the data subject, even though with limited efficiency. Bloom filters can, therefore, only be seen as a 'mitigation' of the risk entailed, but not an 'Avoidance'.

The concept of bloom filters can help to reduce the relatability of a behaviour (e.g. asking for specific transactions) to a data subject. It, therefore, reduces risks associated with disclosed data. It is a PET to minimise data and reduces the risks 'data limitability (minimisation)' (*r_3*), 'unintended disclosure' (*r_4*) and 'unintended access' (*r_8*) would be reduced since the disclosed data would be less relatable to data subjects.

Zero Knowledge Proofs

Zero-knowledge proofs try to solve the issue of the need for a prover to disclose information in order to prove the knowledge to a verifier. The properties of zero-knowledge proofs are [BFM19]:

- *Completeness*, meaning that the verifier will be convinced of the fact;
- *Soundness*, meaning that a verifier cannot be convinced of a lie;
- *Zero-knowledge*, meaning that no verifier learns anything additional than the fact that the statement is true.

Interactive zero-knowledge proofs define a protocol where the prover and verifier communicate about proofing statements that depend on a secret without disclosing the secret itself. Non-Interactive zero-knowledge proofs provide a common random string from the outside to both the prover and the verifier making interaction obsolete [BFM19]. This makes zero-knowledge proofs applicable to many security and PETs. There are many implementations of different Zero-knowledge Proof Systems such as the ‘Sigma Protocol’, Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge (‘zk-SNARK’), Succinct Transparent ARgument of Knowledge (‘STARK’), Zero-Knowledge Prover and Verifier for Boolean Circuits (‘ZKBoo’) or ‘Bulletproofs’. Latter is being used in the implementation of Monero, a privacy-enhancing cryptocurrency. The differences in them lie mainly in runtime, quantum security, proof size and interactivity [Uni21].

In blockchain technologies, these zero-knowledge proofs are used to “prove that a number is found that solves a cryptographic puzzle and fits the hash value without having to reveal the Nonce” [Fen18]. Therefore “it becomes possible to not only perform secret arbitrary computations that are verifiable by anyone but also to do this efficiently” [Rei16]. This is needed to provide a blockchain where the participants know very little about each other, preserving privacy. In the context of data protection, zero-knowledge proofs help nodes not to reveal their identity to each other. This property helps to solve some privacy issues on blockchains stated above. Namely, the protocol-related private data does not solve the problem of private data as payload. It also does not relate to or solve transparency or issues or support the rights of the data subjects.

The concept of zero-knowledge-proofs makes it possible for one party to prove a knowledge of data without having to transmit or disclose the knowledge to another party. The mere information that a party has some knowledge of the data is sufficient in certain use cases and can therefore lead to a lower transfer of (personal) data. ‘unknown transfer in third country’ (r_1) would not apply in case only anonymous data is transferred since zero-knowledge proofs probably only apply to a limited set of data in the exact use case where the knowledge of data needs to be proven it is only reduced. Also, ‘unknown processing activities’ (r_2) would be reduced since less personal data is transferred. It, therefore, reduces the risks ‘Data limitability (minimisation)’ (r_3) and ‘unintended disclosure’ (r_4). It also reduces the risks associated with data breaches ‘unidentified data breach’ (r_6) and ‘unintended access’ (r_8). Also, the ‘right to rectification’, ‘right to erasure’, ‘right to object’ and specifically the risks ‘possible immutability’ (r_9), ‘possible undeletable' (*r_10*) would be reduced if some data does not need to be transferred.

4.2.4 Residual Risk Identification

In every risk management plan, residual risks remain present after treatments have been executed. They may be small, but they should be made aware of. Art. 36 (1) asks for the consultation of the DPA if "processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk" [Cou16a, Art. 36 (1)] and Art. 35 (11) asks for a re-evaluation in case the risk changes, which is arguably the case when reducing the risk by treatment.

In order to manage residual risks, it is important to have an overview and clear understanding of them and to either accept them or treat them (again). As mentioned in the description of the activity 'establishing risk criteria', the effort of the different treatments needs to be weight against the risks to decide whether the risk is acceptable or not.

Table 4.3 summarises the risk treatments and acceptance for public and private blockchains individually.

#	Risk and Impact Description	Risk Priority Private	Acceptance Private	Risk Priority Public	Acceptance Public	Risk Treatment
r_1	Unknown transfer in third country	low	accept	high	treat	Differential Privacy and Noise State channels Ring signatures Homomorphic encryption Hash Functions Off-Chain Identifiers Zero Knowledge Proofs
r_2	Unknown processing activities	low	accept	medium	treat	Differential Privacy and Noise State channels Homomorphic encryption Zero Knowledge Proofs
r_3	Data limitability (minimisation)	medium	treat	high	treat	State channels Onion Routing Pruning Homomorphic encryption Stealth Addresses Bloom Filters Zero Knowledge Proofs

#	Risk and Impact Description	Risk Priority Private	Acceptance Private	Risk Priority Public	Acceptance Public	Risk Treatment
r_4	Unintended disclosure	medium	treat	high	treat	Differential Privacy and Noise State channels Ring signatures Onion Routing Pruning Homomorphic encryption Hash Functions Stealth Addresses Off-Chain Identifiers Encryption Instead of Deletion Bloom Filters Zero Knowledge Proofs
r_6	Unidentified data breach	low	accept	low	accept	Pruning Homomorphic encryption Zero Knowledge Proofs
r_7	Contactability of DPO	low	accept	low	accept	-
r_8	Unintended access	medium	treat	medium	treat	Differential Privacy and Noise State channels Ring signatures Pruning Homomorphic encryption Hash Functions Stealth Addresses Off-Chain Identifiers Encryption Instead of Deletion Bloom Filters Zero Knowledge Proofs

#	**Risk and Impact Description**	**Risk Priority Private**	**Acceptance Private**	**Risk Priority Public**	**Acceptance Public**	**Risk Treatment**
r_9	Possible immutability	high	treat	high	treat	Redactable Blockchain Differential Privacy and Noise Homomorphic encryption Off-Chain Identifiers Encryption Instead of Deletion Zero Knowledge Proofs
r_10	Possible undeletable	high	treat	high	treat	Redactable Blockchain Differential Privacy and Noise Homomorphic encryption Off-Chain Identifiers Encryption Instead of Deletion Zero Knowledge Proofs
r_11	Possibility of recovery	medium	treat	medium	treat	Pruning
r_12	Slow dissemination	medium	treat	medium	treat	State channels
r_13	Contactability of subjects	low	Accept	low	Accept	-
r_14	Readability of format	low	Accept	low	Accept	-
r_15	Unknown data sources	low	Accept	low	Accept	-

Table 4.3: Risk acceptance and treatments for public and private blockchains

The treated risks need to be re-evaluated regarding the residual risks in order to decide whether the solution is sufficient (e.g. another solution is possible and appropriate). Depending on the use case, different treatments are feasible and result in different residual risks.

The risk of unknown transfer in a third country (*r_1*) is acceptable in the case of private Blockchains. On public blockchains, the transfer needs to be treated. The risk treatments can be summarised by either hiding information from Blockchain participants or not giving them access to personal data. Hiding the information by adding security mechanisms such as noise or encryption bears the residual risk of breaking the mechanism. The residual risk also includes hard to predict future advances in technology that might break these mechanisms in the future. Not storing personal data on the blockchain bears the practical risk of accessing the off-chain Data. However, this risk is not higher than in any typical IT system. Therefore, it needs to be tackled with technical and organisational measures like any other IT system. The theoretical risk of legal access to the re-identifying information must also be considered. The immediate risk of transfer cannot be avoided nor mitigated, but the resulting risk for data subjects can be mitigated.

The risk of unknown processing activities (*r_2*) is also acceptable on private blockchains since it is low. The risk with public Blockchains needs to be considered. The treatments can be summarised by either giving only a few parties access to the information, which would allow control of the processing activities or not publishing private data at all. The residual risks are breaking the security mechanisms such as encryption such that another party would be able to process the data transferred. The difference in the risk lies within the communication channel. In the case of state channels, the communication channel is not ephemeral because it is not stored on the blockchain and, therefore, not publicly accessible. This reduces the risk of future attacks on the security mechanisms.

The risk of data limitability (*r_3*) needs to be treated for both public and private blockchains. The solutions can be categorised in limitation of identification of the Blockchain nodes (protocol) and minimisation of Data stored on the blockchain (payload). The removal of identification in the protocol bears the residual risk that the location of the notes will not be identifiable anymore, and therefore processing activities within third countries cannot be identified. Also, other legal obligations such as Anti Money Laundering (AML) and Know Your Customer (KYC) obligations can be negatively influenced. Adding protocol privacy-enhancing features to blockchain technology preserves the identity of nodes, making anonymous payments and illegal activity easier. Anti Money Laundering (AML) regulations, such as European Union's Sixth Anti-Money Laundering Directive (6AMLD or AML6) [Eur18], currently highly depend on identities to track money streams. The AML regulations ask for even tighter KYC processes and AML screenings. Systems that are designed to obfuscate the identity of participants with no central authority and KYC processes in place contradict these regulations. Removing unneeded information at the nearest possible time, on the other hand (pruning), is nearly risk-free. Adding a layer of communication above the Blockchain layer, such as state channels, mix the minimisation easier since an active request and response allows for the transfer of only needed information. Therefore it should be discussed which information needs to be permanently stored on a Blockchain and which is only needed for a short period. The residual risk with all treatments implemented heavily depends on the implementation of these technologies.

The risk of unintended disclosure (r_4) needs to be treated on public and private Blockchains. The risk is system imminent since all communication information is publicly available. The treatments can be categorised as either hiding information, not transferring it or transferring it over different, less open channels. The risk of unintended disclosure is probably the one with the most impact on data subjects. Hiding the information with a residual risk that is very hard to estimate since it relies on technological advances in the future is not sufficient. Treatment must be highly independent of future advances to reduce the residual risk to an acceptable minimum.

The risk of identified data breaches (r_6) was acceptable for both public and private blockchains. However, the risk might be reduced by pruning and encryption mechanisms. The risk of contact ability of DPO (r_7) was accepted in both cases private and public blockchains. The risk of unintended access (r_8) needs to be treated for private and public blockchains. The access to transferred or stored information (confidentiality) is, on the one hand, existent in protocol information and, on the other hand, in the payload. Different security mechanisms, such as encryption and authorisation, help mitigate the risk within the treatments. In the case of long-time stored information, these treatments bear the residual risks of broken encryption in the future. In the case of deletion, the residual risk exists that data could be recovered.

The risks of possible immutability (r_9) and possible and deletable (r_10) are often discussed in private and public blockchain issues. The risks indirectly influence other risks since other risks might be reduced if these risks are avoided. Since immutability is one of the cornerstones of the data structure of Blockchain technology, it exists in both private and public Blockchains. The solutions can be categorised as either changing the technology completely (retractable blockchain) or avoiding personal data at all, or storing information off-chain. In the case of retractable blockchain, the residual risk is that in the case of changeable or deletable information, the Blockchain concept might break, meaning that information can be altered by political interests or censored. Storing encryption keys off-chain and deleting these keys instead of deleting the data from the blockchain bears the residual risk that the encryption might break in the future, and data would still be accessible.

The risk possibility of recovery (r_11) needs to be treated for both public and private blockchains. Treatment against the recovery of information is securely destroying or deleting data. The residual risk depends on how secure the Data is deleted.

The risk of slow dissemination (r_12) is imminent in distributed non-real-time systems. Therefore the treatment found in the literature is to have direct communication channels (e.g. state channels). The risk can be reduced to a low probability.

The risks of contactability of subjects (*r_13*), readability of format (*r_14*), and unknown data sources (*r_15*) are accepted for both private and public Blockchains.

4.3 Discussion

The GDPR applies to all applications operated for the processing of personal data by controllers or processors within the EU. Blockchains are distributed applications that process data in the protocol and payload. The specific characteristics of blockchains pose risks to compliance with the GDPR.

In this chapter, a structured risk assessment process based on the GDPR has been developed to identify and assess risks. Risk treatments for the risks found were researched in the current literature, and the remaining residual risks were qualified.

First, it was shown that personal data is present in the protocol and payload of blockchains and is processed by legal entities inside and outside the EU. In Section 4.2.2 it was shown that these data could relate to EU citizens, and since public blockchains are generally accessible to anyone on the internet, including to the European Union, the service is offered to European citizens. All these facts require regulation under the GDPR.

The research has answered the question "How does a public blockchain change GDPR compliance in an application?" by providing a structured analysis of obligations and rights under the GDPR and applying this to technology. The problems with using blockchain in a privacy-compliant way were described. In addition, other possible solutions related to GDPR compliance and data protection issues in the literature were discussed, and conclusions were drawn.

Previous works mentioned in the analysis either focus on the technical implementation of achieving more privacy in the protocol and payload or they focus on the legal aspects of blockchain as a technology. As the problem is closely linked to a specific EU legislation, namely the General Data Protection Regulation, and the technology is extensive and rapidly progressing, it is difficult for both sides to stay abreast.

Current data protection laws legally define roles that can hardly be applied to decentralised authorities. The roles discussed in the literature, such as that of the originator of a decentralised system, are not included in the legislation. Interpreting shared control to mean that each participant in a blockchain network has the power to define means and ends leads to an unpredictable and probably unenforceable level of regulation.

The research shows that there are many GDPR-related issues and risks with blockchains. It also shows that there are technical solutions to many problems. The choice of the solution must be made according to the use case, as solutions not only minimise risks but can also negatively influence the use case. The approach of the risk analysis presented can also be applied to other technologies, and the result of the analysis can help application developers and architects to decide whether to use blockchain technology in a GDPR-sensitive environment at all and which technologies they should use to protect data subjects.

The study conducted is a snapshot of technology and legal opinion and can change within a very short period of time. Indeed, it is likely that technological advances will continue to develop at the pace described in the near future. Nevertheless, the study helps to understand the current state of the art and its applicability in solving problems related to the GDPR.

In general, the research area of data protection, which finds its application in this thesis between law and computer science, is exposed to enormous changes due to technological progress and new legal regulations. Data protection risks resulting from the use of blockchain technology can be reduced by technical means and technologies but cannot be completely excluded. Therefore, it is recommended to reconsider the use of blockchain technology for the direct storage of personal data, even if blockchain technology is not used for this purpose.

CHAPTER 5

Design of an Evaluation Process for Legal Process Supporting Technologies

The previous chapters have shown the fundamentals and risks in the fields of data protection and blockchain technology. It was shown that certain aspects of data protection according to the GDPR are in conflict with the fundamental properties of blockchains. However, the influence between blockchain technology and data protection also extends to the supporting features of blockchain technology to actively benefit data protection and the right and freedoms of data subjects. Which problems exist and to what extent they could be solved with the help of blockchain technology is questionable. The scientific research questions defined in Chapter 1 therefore are accordingly "How can blockchain technology be used in applications to support data protection according to the GDPR?" (Question 2.1) and "How can blockchain-based applications to support data protection according to the GDPR help empower data subjects?" (Question 2.2).

To assess whether blockchain technology can improve aspects of data protection in some instances, the Design Science Research Methodology (DSRM) was taken to scientifically answer the research question.

The research question of using technology to support privacy in terms of data-protected applications is a broad field. Privacy Enhancing Technologies (PETs) seeks to improve this area. These technologies focus on finding implementations in the organisational environment that enable privacy-compliant implementations. PETs help to implement high privacy standards.

Technologies, in general, as the application of scientific knowledge, try to transfer scientific advancements into engineering. The field of Legal Technologies (Legal Tech) applies these technologies to enhance the lawyer's "ability to perform physical or intellectual tasks" [WFMC17, p. 22ff].

Privacy Enhancing Technology (PET) try "protecting informational privacy by eliminating or minimising personal data, thereby preventing unnecessary or unwanted processing of personal data" [BVE$^+$03].

[PTMT19] define goals and targets of Privacy Enhancing Technologies (PETs) broader as followed:

- Communication protection: Security, Anonymity
- Data protection: Integrity, Confidentiality
- Entity authentication: Identity-based, Attribute-based
- Privacy-aware computation: Confidential inputs, Privacy-adding
- Human–data interaction: Transparency of data usage, Intervenability

They used the terms from the ENISA defined in [DDFH$^+$15]. [RP09] already coined the privacy protection goals unlinkability, transparency, and intervenability.

'Unlinkability' means that data is not linkable outside the domain it is used in. 'Transparency' means that "the legal, technical and organisational setting can be understood and reconstructed at any time" [DDFH$^+$15], meaning before and after the processing operation. 'Intervenability' means that the data subject can intervene in the processing of his data.

'Unlinkability' has been discussed in the previous chapter focusing on blockchain technology to solve privacy issues. Many technologies such as onion routing, differential privacy and bloom filters exist to advance unlinkability.

'Transparency' has been discussed as, on the one hand, privacy reducing, but on the other hand, rights enhancing feature in the means of blockchain technology. The transparency of blockchain computations and data helps data subjects understand the processing activities and access their data. It, on the other hand, allows others to process their data, reducing their right to decide who processes their data.

'Intervenability' is clearly defined by data protection legislation but is little worked on by PET. Rights-supporting technologies could help data subjects to intervene in data processing activities and help empower data subjects and improve data protection from the perspective of data subjects.

The applications to be investigated can be classified as intervention and transparency technologies, as sub-areas of technologies and applications that support natural and legal persons in exercising their legally defined rights on the one hand and obligations on the other.

To allow to evaluate applications regarding their supporting features on the basis of criteria based on a problem statement found via quantification of issues in the processes extracted from legal texts, the following approach was designed on the basis of the steps defined by DSRM (see Figure 5.1).

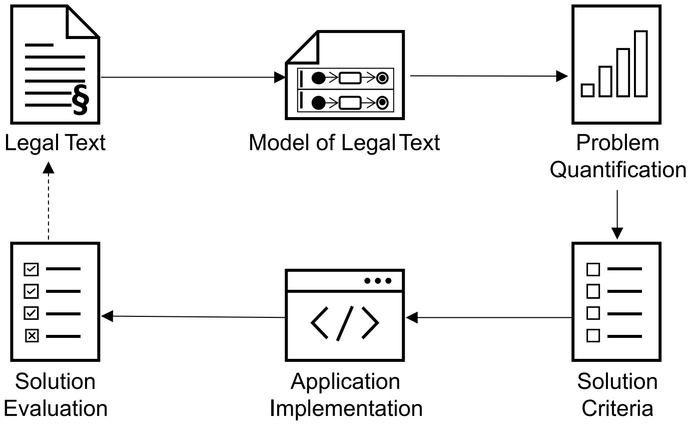


Figure 5.1: Possible circular process to evaluate applications on the basis of criteria based on a problem statement found via quantification of issues in the processes of legal texts

The process starts by analysing a legal text. Roles, activities and decisions are extracted. These are processed into a model that describes the process in the legal text ('Model of Legal Text'). This process is then used to make a targeted analysis of which process steps are perceived as difficult ('Problem Quantification'). Based on the generated artefacts, structured criteria for solutions in the individual process steps are developed ('Solution Criteria'). An existing or newly developed application can be compared with these criteria ('Application Implementation'). For this purpose, a criteria catalogue with criteria questions is generated. These serve to evaluate the application ('Solution Evaluation').

With this process, it is possible to scientifically find and solve issues in legal texts. The following sections describe the process in greater detail.

5.1 Modelling of Legal Text

The positive law in legal texts entails authority over affected persons. It generates legal consequences that can have real effects on a person's life and freedom. This legal text contains norms that are codified by prohibitions, commandments and optional provisions. A mapping of the legal texts and the logic hidden in them by analysing the texts is attempted by science. Science is also striving to create an automated interpretation and modelling of legal texts in order to create legal support systems that can assist in legal decisions.

[ISS$^+$13, p. 69ff] used the 'Nòmos' framework to build their model. They designed a simple three-step model: Identify the legal roles, the situations, legal operators and relations. [CVG$^+$15, p. 1401ff] designed a similar modelling technique, the 'Structured Process Modeling Theory'. [CWVK11,] also investigated several ways to turn laws into models. [AMMS19] describe the modelling of GDPR processes, with cases from a phone company focusing on the compliance of the data controller. They created design patterns to be used in Business Process Model and Notation (BPMN) for implementing privacy-enhancing features according to the GDPR.

Business Process Model and Notation (BPMN) has been used to model requirements and complex systems. It is a graphical specification language for describing processes, choreographies and collaborations [OMG14] and was specified by the Object Management Group (OMG). BPMN 2.0 is a recognised standard for modelling business processes with a comprehensive set of symbols for business process diagrams.

Basic elements of a BPMN are activities, which are an abstract representation of an operation of a system or a person; connections, which represent the sequence flow of activities; gateways, which represent decisions in the sequence flow, such as whether activities are executed in parallel, or only one sequence flow; actors or systems, which are represented by swimlanes.

Modelling laws with BPMN has been researched and practically used. [CWVK11,] formulated the idea in 2011 that BPMN could be used to model laws for the Italian immigration law in order to make it more understandable. Their approach added special markup to the laws in order to produce process models.

There are approaches [BCM19a, DMMRP19, BF20] for modelling applications in BPMN and adding metadata to steps in order to check for compliance of the application, but not modelling laws to check for compliance.

[HM20] describe how process analysis can be used in system engineering. Their BPMN process for rights according to the GDPR is very general and limited, namely six tasks for all cases. They specify tasks in the general software development process with the use of PETs to create privacy-aware systems. They define work units and work products to be created during the design and implementation of a system.

[BCM19b] suggest the extension of BPMN to model processing activities according to GDPR within a BPM to transfer knowledge about processing activities using the 'DAPRECO' model. It describes creating a structured representation of the legal text, a conceptual model of the terms and a machine-readable translation of the provisions. The result is a BPMN diagram annotated with processing properties such as type of processing and properties defined by the type, for example, transfer to a third country, whereas the country is a property. [PTMT19] used PE-BPMN to define computing processes with a detailed description of privacy-enhancing technologies used. They describe computing processes more than business or real-world processes in order to be able to use them according to security or privacy technologies. [BHLR12] focus on the access control part of privacy enhancement and define an extension of BPMN called 'SecureBPMN'. This leads to a traceable security model for access control defined on a business level rather than a technological level, helping to have an end-to-end discussion of access control mechanisms. [SDG14] describe an extension of BPMN to check security policies against BPMs. They define modelling language 'SecBPMN' to annotate activities with predicates such as ‘accountability’, ‘integrity’ or ‘reproducibility’. A query language, ‘SecBPMN-Q’, is defined to model security policies using security annotations.

These models are an informative representation to show a certain aspect of a process to a certain group of readers. In the case of rights-based applications, the processes are modelled from the perspective of the affected parties. The process of the individual is in the foreground.

A model of a legal text can then be used to structure analysis and quantify problems in the activities. Furthermore, it can be used to create qualitative factors to support the process steps.

5.2 Problem Quantification

The problem quantification can be done in several ways. The result must be some kind of understanding of where the problems lie within the model described and how severe the problem is. The objective of the empirical research is to find out which process steps are considered difficult to carry out by the population. A qualitative or quantitative approach can be taken, depending on the difficulty of the process and the size of the population researched. The resulting hypotheses that needed to be tested in the course of the research are as follows:

Hypothesis 1: Process steps are currently perceived as easy to carry out by those who have carried them out.

Hypothesis 0: Process steps have been perceived as difficult to carry out by those who have carried them out.

The problem quantification helps to assure that issues exist and to focus on certain steps that have issues. It can additionally be used to create the basis for the next step by qualifying certain issues with the steps.

5.3 Solution Criteria

The solution criteria is a qualitative representation of the factors assumed to lead to a higher quality of execution of an activity. The main question to be answered within this step is “How can the solutions supporting certain activities in the process be evaluated?”. The criteria for each activity in the BPMN can then be added to the model as an annotation. [BHLR12] extended BPMN in order to model security criteria. [SDG14] defined a modelling language ‘SecBPMN’ to annotate activities with predicates such as ‘accountability, ‘integrity’ or ‘reproducibility’. A similar approach can be taken to describe and evaluate criteria for certain process steps.

The scientific approach to finding solution criteria must be a creative process with an open solution space. A quantitative approach, as used in the problem quantification, can therefore not be used since it is very much structured and gives little space for creativity [Hie09, p. 116].

Qualitative empirical research, in contrast, is intended to create an unbiased recording of reality, representing existing as well as possibly newly emerged complex constellations and mechanisms found in the field of investigation [Fli95]. The quality of qualitative research needs to be assured by setting quality standards for the process [Lam16], by implementing Mayring’s six criteria [May16] during the qualitative research:

- procedural documentation;
- securing interpretation with arguments;
- rule-guidedness;
- proximity to the subject;
- communicative validation and triangulation.

This particular research can be done with expert interviews, focus groups, the ‘Delphi’ method [Bro68], or structured brainstorming sessions [BB93]. During this process, each activity needs to be analysed with regard to steps taken and inputs needed. The quality factors of the activity are trying to be evaluated by asking these questions. In the case of using an interview technique, the questions need to be formulated as given by qualitative guidelines [Str19].

In order to obtain as diverse information as possible, the participating individuals should contribute different perspectives. Participants in different positions in a problem-solving process tend to have different views. In the later course, the criteria of the sample can be adjusted and circularly processed until it leads to a theoretical saturation [DB16, p. 302].

In any case, the conduct of qualitative research must be recorded. A common method is audio or video recording and subsequent transcription. The transcription then needs to be analysed. A common method of analysing qualitative interviews is Braun and Clarke’s [BC06] ‘Thematic Text Analysis’ procedure. It defines six steps to generate central themes [BC06, pp. 15-24]. These themes are used to summarise properties needed in the according activity and to generate corresponding criteria (see Step 2 in Figure 5.2). These criteria are structured using evaluation fields and are always linked to a specific BPMN activity. This enables a structured, process-based evaluation.

These criteria can, on the one hand, be used for the generation of requirements for implementation and, on the other hand, for the evaluation of applications.

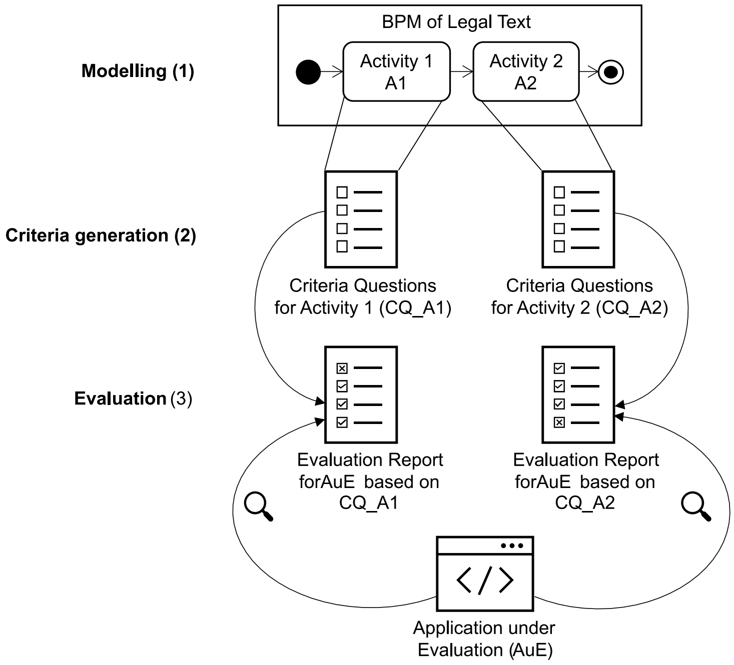

BPM of Legal Text
Activity 1 A1
Activity 2 A2
Modelling (1)
Criteria generation (2)
Criteria Questions for Activity 1 (CQ_A1)
Criteria Questions for Activity 2 (CQ_A2)
Evaluation (3)
Evaluation Report forAuE based on CQ_A1
Evaluation Report forAuE based on CQ_A2
Application under Evaluation (AuE)

Figure 5.2: Study design and evaluation on the basis of a business process model of a legal text

5.4 Implementation

Solutions regarding the evaluated process activities might be any rights-supporting technology. It might be any artifact that helps the activity. This can, for example, be a process, software, hardware or tool. The implementation of the rights supporting technology can be done with any structured development process. The technology itself is not defined by the solution criteria but can be judged upon the criteria. The model of the legal text, the problem quantification and the solution criteria can be used as a basis for the technical specification. Requirements are then defined as capabilities the system must meet to satisfy the specification [IEE90]. The technology can implement process steps of the created model or certain aspects of the quantified problems. In the development of the model of the legal text, several stakeholders were identified. Since the model was created from the perspective of certain affected groups, these groups can be used within the development process. These include but are not limited to usability and acceptance tests.

5.5 Solution Evaluation

A solution, whether implemented on the basis of solution criteria or not, can be evaluated to fulfil the solution criteria. The solution evaluation creates an evaluation report from the 'Application under Evaluation' (see Step 3 of in Figure 5.2). This report is generated by the evaluation of the defined solution criteria questions.

Since the criteria and, subsequently, the questions are directly related to activities, the report can be visualised with the help of the model representation. The evaluation then shows the support for each activity within the model.

5.6 Discussion

This chapter designed a process on the basis of the scientific Design Science Research Methodology (DSRM) to evaluate legal supporting technologies by modelling processes defined in a legal text, quantifying problems, finding solution criteria, and evaluating applications. It ultimately indicates how strongly an application supports activities in a process. It can be used to design applications and evaluate them.

In the future, it could be examined whether this process can also create a feedback loop and a positive influence can be carried out on the basis of the findings of the processes of those affected.

The designed process is now applied to a real-world scenario, namely the erasure and access process defined by the General Data Protection Regulation (GDPR) and two applications helping data subjects to execute the processes defined by the law.

CHAPTER 6

Use of Blockchain Technology to Support Data Subject Rights

Blockchain technology is used for many purposes (see Section 3.5). This can presumably also happen for the purpose of data protection. This chapter explores the use of blockchain-based technology in data protection applications.

For this purpose, the process presented in Chapter 5 is applied in order to be able to make a scientific statement.

First, two aspects of the GDPR are modelled, the right to be forgotten and the right to access. Both are modelled in BPMN from a global perspective but primarily from the perspective of the data subject. Afterwards, any problems in the processes are identified and quantified through a quantitative survey. The same process is taken to experts, and those activities that have revealed problems in the survey are subjected to a structured solution process. For this purpose, evaluation criteria for solutions are created by interviewing experts, application prototypes are evaluated, and conclusions are drawn from the results.

6.1 Model of GDPR Processes

The properties of blockchain technologies described in Section 3.2.1 help to develop applications with transparency and immutability features. They might be used to allow for data protection enhancements and rights-supporting applications.

To assess these applications, processes of the GDPR were modelled in order to gain insights into the roles, activities and dependencies. A standard BPMN model was created for a structured definition of the case studies. The model shows two core processes and core roles as defined in Section (2.3). The following processes are models of the most important GDPR processes.

Modeling laws, as described in Section 5.1, is a common practice in science. For example, the GDPR was modeled by [HM20] and [AMMS19].

[AMMS19] describe the modelling of GDPR processes, with cases from a telephone company, focusing on the compliance of the data controller. Based on the model, they created design patterns that can be used in Business Process Model and Notation (BPMN) for the implementation of privacy-friendly functions according to the GDPR.

[HM20] also describe how process analysis can be used in system engineering in general. Their BPMN process for rights under the GDPR is very general and of limited use. The model consists of six activities and does not take into account special cases of different rights exercises.

Both processes have been taken into account while modelling the processes within this work. The difference between the models created in this dissertation and those in the literature is that the modelling in this dissertation has been done with a data subject orientation because of the objective to enhance the data subject's data privacy. This means that the processes were modelled from the perspective of a citizen who wants to exercise his or her rights.

6.1.1 Process of Deletion after Request for Erasure by Data Subject according to the GDPR

The usual process of a processing activity is shown in Figure 6.1). It represents the usual process of a processing activity of personal data by a controller with the help of a service provider (for example, a print service). The process starts with the disclosure of personal information to the service provider, which is in terms of the GDPR usually the controller (see Section 2.3.1). The controller processes the data itself and eventually stores it in a database. It also sends the data to a subcontractor, usually a process within the scope of the GDPR. The processor also processes the personal data and eventually stores it in a database. After the legal basis or purpose for which the data was stored has expired, the data must be deleted.

DIN 66398 describes the necessity for a deletion concept. Even though it was defined before GDPR, it describes concepts that should be considered. [HM20] interpret [Ham16] and find the following elements needed in a deletion concept:

1. Deletion rules: 'Deletion classes' are defined for combinations of holding periods and starting times. These holding periods could either follow directly from the legal provisions or could be defined by the company. For each deletion class, a deletion rule is specified.
2. Implementation instructions: The technology-agnostic standard deletion rules are detailed in the implementation guidelines.
3. Exceptions: To allow for necessary flexibility, e.g. in case of lawsuits, exception rules can be defined.

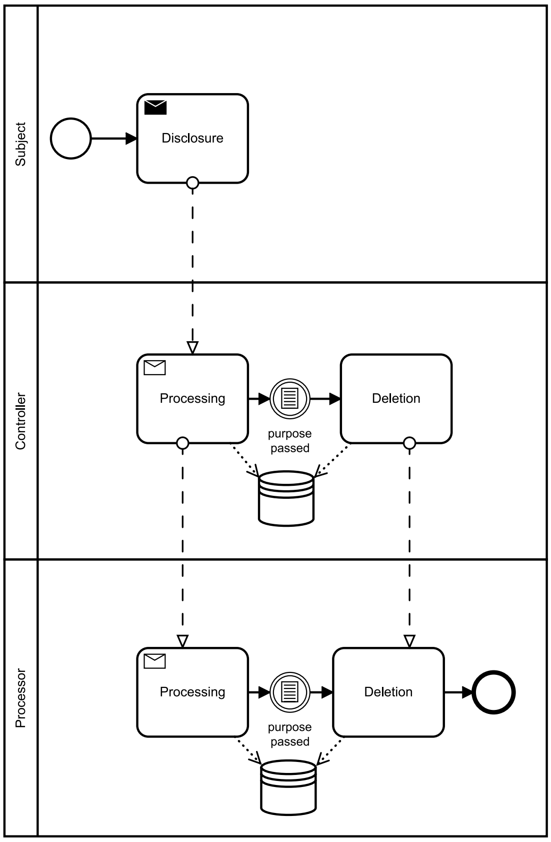


Figure 6.1: Usual minimal process of a processing activity

4. Documentation: Deletion rules, implementation instructions and exceptions should be stored separately.
5. Responsibilities: The different stakeholders of the deletion concept must be assigned to the tasks that are specified by the norm.

An erasure request is an informal request to a controller or processor to remove personal data of a data subject in accordance with [Cou16a, Art. 17]. The request for erasure expresses a data subject's willingness for the controller to remove all personal data or identifiable characteristics from all processed data sets. Hence, to delete or anonymise the personal data. In the following sections, this process is modelled with a focus on data subjects and controllers.

Model of the Erasure Process from a Data Subject's Perspective

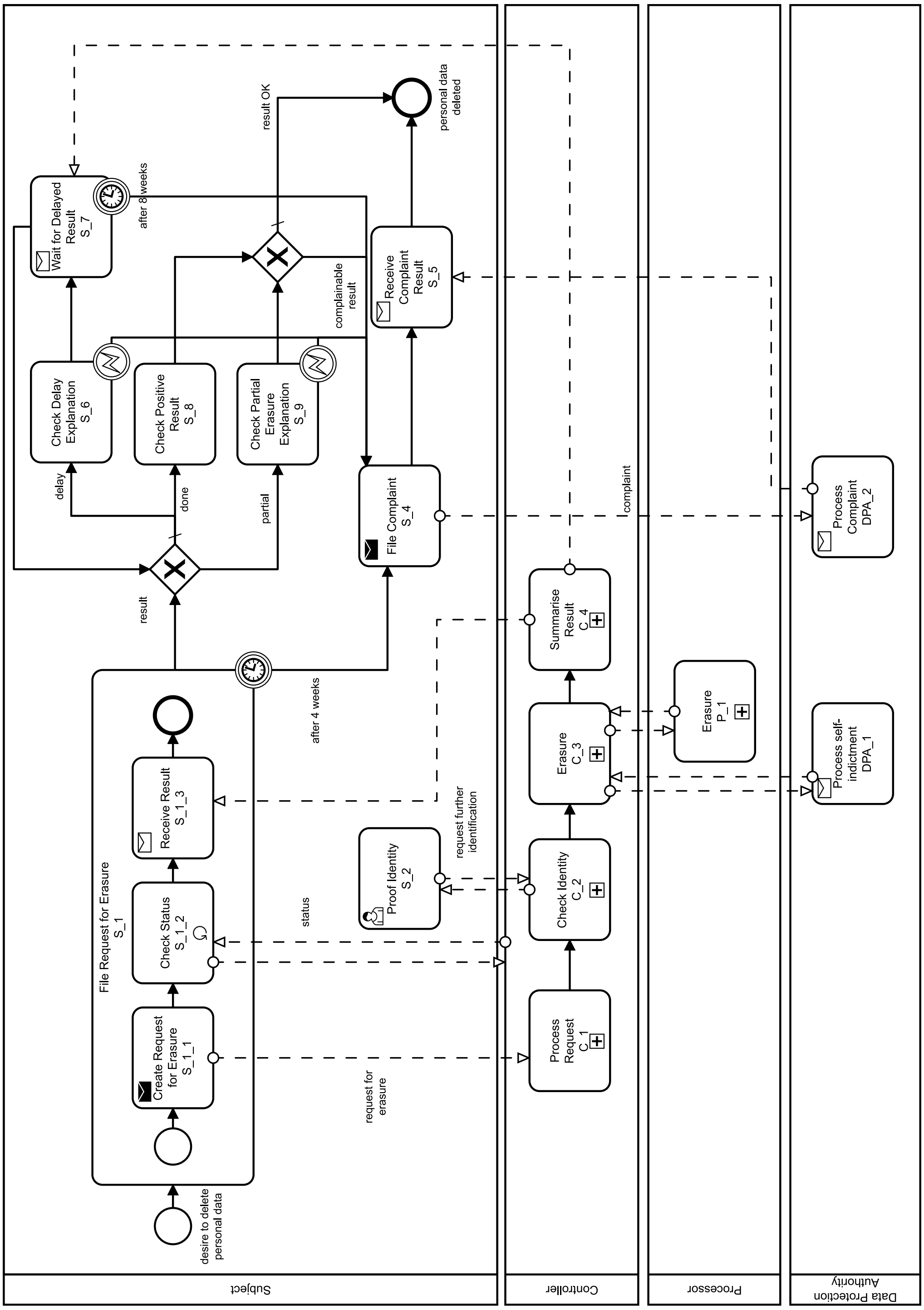


Figure 6.2: Erasure process focused on the data subject

Figure 6.2 shows the process model of the erasure process with a focus on the data subject. Activities with a subprocess sign (plus sign) are detailed in another Figure. A full representation of the processes can be found in Appendix B. The process of a 'request for erasure' starts with a request to the controller (*S_1_1*). This request can be made via any form of communication in an informal way, meaning that there is no need to use a specific form, and the controller may not require any specific form. In case the controller cannot identify the data subject, it can ask for additional proof of identity (*S_2*) [Cou16a, Art. 12 (2)]. The controller shall use "all reasonable measures to verify the identity of a data subject who requests access, in particular in the context of online services and online identifiers"[Cou16a, Rec. 64] according to Recital 64. The controller, though, is not "obliged to acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of this Regulation"[Cou16a, Rec. 57] according to Recital 57. This can be done in different ways. The most common method at present is to ask the applicant for a scan or picture of their identity card or passport, login or customer identification. The controller receives and processes the request (*C_1*). It usually checks the database for past GDPR requests and whether personal data are stored in one or more systems. Depending on the architecture and the use of privacy-friendly software, this can take days to weeks.

After a result has been received (*S_7*), it is checked (*S_8*). The data subject might ask for the status of the request if it takes longer (*S_1_2*). If the request has not been answered (*S_1_3*) within four weeks, a complaint can be filed with the data protection authority (*S_4*). A request can also be answered by reasoning why the processing of the request might take longer than four weeks with a maximum of twelve weeks (delay) [Cou16a, Art. 12]. If the explanation of the delay is not sufficient or not understandable (*S_6*), the data subject can again file a complaint. Also, a partial erasure can occur (*S_9*) if the controller can argue why it needs parts of the data (see next section) with (another) legal basis. The complaint (*S_4*) can be made via an email with a PDF form (an example form can be found at the Austrian Data Protection Authority website[1]). This form again requires a process such as identification and proof of the request (*DPA_1*). In the case of a formal complaint, the data subject receives a result from the DPA (*S_5*).

Ideally, however, a complaint is not necessary, and the controller deletes the data (*C_3*) and checks if the personal data has been sent to an external processor (*C_4*). If so, it forwards the deletion request to the respective processors, which consequently execute the deletion (*P_1*). This process could have more involved sub-processors (not depicted in the model).

Model of the Erasure Process from a Controller's Perspective

[1] https://www.dsb.gv.at/download-links/dokumente.html Accessed: 23.04.2021

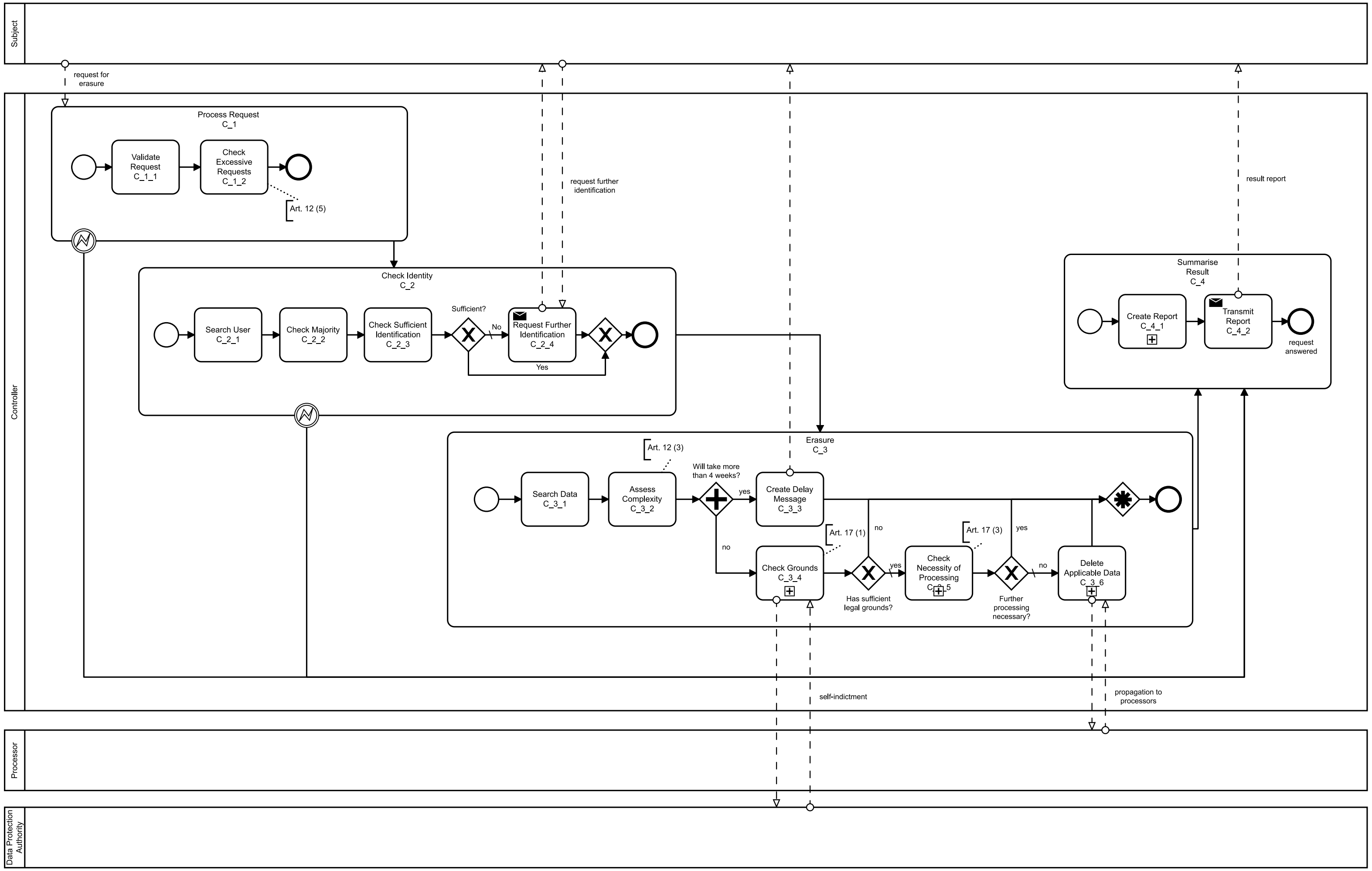


Figure 6.3: Erasure process focused on the controller

Figure 6.3 shows the process model of the erasure process from the perspective of a data subject. Activities with a subprocess sign (plus sign) are detailed in another Figure. A full representation of the processes can be found in Appendix B. The figure shows the data subject in a collapsed form as a sender of messages. The first initial activity started by a message by the data subject is the processing of the request (*C_1*). Within this process, the controller first validates the request (*C_1_1*). It checks whether all the necessary information has been provided and whether the application is founded [Cou16a, Art. 12 (5)]. If this is not the case, the exception (lightning symbol) is executed, which sends an error report to the data subject. After that, it is checked if the same or an equivalent request has been made excessively (*C_1_2*). If this is the case, the controller has the right to charge a fee [Cou16a, Art. 12 (5) lit. a] (not depicted) or to reject the request [Cou16a, Art. 12 (5) lit. b]. After that, the check identity activity is done (*C_2*). First, the controller searches for the data subject's identity in its database (*C_2_1*). If the user is not found, a respective summary result is transmitted to the data subject. After that, the majority of the data subject is checked (*C_2_2*) based on the identity. If this is the case, an authorisation is needed from the holder of parental responsibility. After that, it is checked if the authorisation for the request for erasure is given, e.g. the identity is sufficiently verified (*C_2_3*). If this is not the case, further identification information is requested from the data subject (*C_2_4*). This concludes the activity of establishing identity.

After that, the actual Erasure activity starts (*C_3*). This process starts by searching for data related to the data subject (*C_3_1*). In the case of distributed or complex systems, this can be a complicated, time-consuming task. When knowing the complexity of the task, it has to be checked if the execution of the erasure will take more than four weeks (*C_3_2*). If the execution takes more than four weeks, [Cou16a, Art. 12 (3)] demands a message to the data subject about the complexity of the erasure and the delay (*C_3_3*). The extension of the execution time can be a maximum of two months [Cou16a, Art. 12 (3)]. In parallel, the legal grounds for the request for erasure are checked. Figure B.1 shows the details of the subprocess 'check grounds' (*C_3_4*). First, the stored legal basis on which the data has been processed needs to be recalled (*C_3_4_1*). This is needed in order to decide whether further processing on the basis of another legal basis can be done and if a withdrawal of consent is possible. The latter is only the case if the legal basis was a consent of the data subject (or the holder of parental responsibility). For this reason, also the purposes of the processing activities and the age of the data subject are needed (*C_3_4_2* and *C_3_4_3*). Article 17 (1) describes different grounds that might be chosen by the data subject. In case the ground was according to Art. 17 (1) lit. a the controller checks if the personal data is still needed for the given purpose (*C_3_4_5*). In case the ground was according to Art. 17 (1) lit. b the controller checks if the basis was according to Art. 6 (1) lit. a or Art. 9 (2) lit. a and if so, no other legal ground for further processing exists (*C_3_4_6*). In case the ground was according to Art. 17 (1) lit. c the controller must check whether the purpose was for direct marketing [Cou16a, Art. 21 (2)] or whether the controller has 'compelling legitimate grounds' [Cou16a, Art. 21 (1)] or a legal case that needs the data [Cou16a, Art. 21 (1)] (*C_3_4_7*). In case the ground was, according to Art. 17 (1) lit. d the controller checks the reasoning behind the unlawfulness of the processing and eventually self-indictment at the Data Protection Authority (DPA) (*C_3_4_8*). In case the ground was, according to Art. 17 (1) lit. e the controller checks the (EU or local) law if it constitutes a legal obligation to delete the data (*C_3_4_9*). In case the ground was, according to Art. 17 (1) lit. f the age of the data subject needs to be checked and whether the consent given by the holder of parental responsibility is valid (*C_3_4_10*). Otherwise, the data needs to be deleted.

If the request for erasure has sufficient legal grounds, the controller checks the necessity of processing (*C_3_5*). This could be the case if the right to freedom of expression applies according to Article 10 of the Human Rights Act applies (*C_3_5_1*) [Cou16a, Art. 17 (3) lit. a]. If this does not apply, the controller can check if other legal obligations exist (*C_3_5_2*), such as taxes or anti-money laundry obligations or if a public interest exists that is higher than the interest of the data subject (*C_3_5_3*) or if historical research purposes or statistical purposes in accordance with Art. 89 (1) apply (*C_3_5_4*). Lastly, the existence of any legal claims against the controller or any defence against legal claims is foreseeable (*C_3_5_5*) [Cou16a, Art. 17 (3) lit. e].

If the legal ground is sufficient and no legal basis for further processing exists, the actual erasure can proceed (*C_3_6*). First, the personal identifiers of the legal person (data subject) need to be collected (*P_1_1*); then, personal data needs to be found in backup and live production data (*P_1_2* and *P_1_3*). Any data that cannot be deleted needs to be marked so that it cannot be further processed in the future (*P_1_4*). This can be the case if an immutable backup. According to CNIL the backup may remain, but this is automatically deleted after a certain time (retention period), but in case of a recovery, the data of the data subject may not be restored or must be deleted immediately [ndledl20]. Other data needs to be deleted or, if possible, anonymised (*P_1_5*).

After that, a report is generated, either with the reasons why the personal data is not deleted or a deletion confirmation. Both must be done using clear and plain language [Cou16a, Art. 12 (1)].

6.1.2 Process of Data Retrieval with a Request for Access by the Data Subject according to the GDPR

A data subject or perceived data subject has the right to information under [Cou16a, Art. 15]. The right of access is according to [Cou16a, Art. 15 (1)] a data subject's right to obtain from the controller confirmation as to whether or not personal data concerning her or him are being processed and, where that is the case, access to the personal data and information specified in [Cou16a, Art. 15 (1)]. This includes purposes of processing activities; categories, source, recipients, retention time of personal data, and information of rights and whether automated decision-making or profiling is done. [Cou16a, Art. 15 (3)] also grants the right to request a copy of the data processed. Recital 63 also extends to the automatic processing to the "logic involved in any automatic personal data processing" [Cou16a, Rec. 63] and the consequences of profiling.

The right to data portability [Cou16a, Art. 20] also allows the data subject to request its personal data. It covers personal data the data subject has provided to the controller [Cou16a, Art. 20]. The data subject has the right to receive these data in a structured, common, machine-readable format [Cou16a, Art. 20 (1)] and has the right to ask for a transfer to another controller were feasible [Cou16a, Art. 20 (2)]. This is subject to the condition that the processing is carried out either on the basis of consent or explicit authorisation concerning specific categories of personal data or the performance of a contract. The right of portability is not provided for in the case of processing for other reasons or legal grounds other than consent or contract, nor for data which have not been collected from the data subject.

Model of the Request for Access from a Data Subject's Perspective

The process request for access starts with creating a request (S_1_1). This request contains the information needed for the processor to gather the information and send it back to the data subject. A standard PDF form can be found on the webpage of the data protection authority of Austria.

If the controller cannot identify the data subject, the controller is obligated to ask for further proof of identity by the data subject (S_2).

Often controllers have a platform or a web service to request the data and see the status of the request. The controller does not have a factual obligation to deliver the status of the request. The GDPR only obligates the controller to notify the data subject if the request takes more than four weeks in total. The controller can notify the data subject that the request will take longer with a maximum extension of 3 months (S_6). The data subject receives a result via the channel he or she requested access (S_1_3). If either a rejection was not based on a legal explanation (S_8) or the request has been delayed for more than four weeks without an explanation, or the provided information is not sufficient (S_9), the data subject may file an official complaint at the data protection authority (S_4). The data protection authority may answer the official complaint within a reasonable time (Rec. 141) (S_5).

Model of the Request for Access from a Data Controller's Perspective

The controller receives the request and validates the request (C_1_1) by checking if all needed information is filled out and the request can be fulfilled. After that, the controller checks whether the data subject has requested his or her information multiple times (C_1_2) and decides whether it asks for financial compensation (see Art. 12 (5)) or rejects the request. The controller then searches for the data subject (C_2_1) and tries to check the age of the data subject (C_2_2) to figure out whether the data subject is of age. This is needed since a minor needs the authority of his legal guardian or holder of parental responsibility. Then the controller checks whether a clear identification of the data subject is possible (C_2_3). If this is not possible, the controller shall request further identification from the data subject (C_2_4). Then the controller searches for the whereabouts of the data of the data subject (*C_3_1*) to assess the complexity of the request (*C_3_2*). If the data is spread to multiple systems, the retrieval might take longer. If the answer to the request will take more than four weeks to retrieve, the data subject receives a delay message (*C_3_3*). In parallel, the collection of the processing information starts (*C_3_4*). The following information needs to be collected by the controller: The purposes of the processing (*C_3_5_1*); the categories of personal data being processed (*C_3_5_2*); if the recipients of the data, especially if they are in a third country (*C_3_5_3*); the storage period and retention time of data stored by the controller (*C_3_5_4*); the source of data not directly collected from the data subject (*C_3_5_5*); an explanation about automated decision-making and profiling and what consequences the decision-making has on the data subject (*C_3_5_6*); if data is transferred to third countries explanation of safeguards on the data being transferred (*C_3_5_6*). After that, the controller must collect the data subject's rights and explain them in the final report (*C_3_5*). Then the controller collects the actual data from the data basis and might ask the processors of data for additional data stored with them (*C_3_6*). Then the controller compresses the information in an application archive and uploads it to a system accessible by the data subject (usually directly via mail or a web service if too large). In any case, the results are summarised, and a report is created (*C_4_1*) and transmitted to the data subject (*C_4_2*).

6.2 Quantitative Discovery and Evaluation of the Problems

It must be determined whether or not there are problems with the request for erasure according to the GDPR. In addition, more detailed insight into the problems must be gained in order to understand and, if possible, solve them.

6.2.1 Design

The identification and elaboration of issues were quantified with quantitative research. The quantitative method enables a precise verification of the hypotheses compared to qualitative methods [Sti05, p. 91]. The quantitative study allows a high number of participants to be interviewed because of the small effort of repentance. For this quantitative study, an interview in the form of a questionnaire was used. Since the topic may include admissions of weaknesses and failures, an anonymous question-based survey was chosen because more honest answers can be expected [May09, p. 100]. A disadvantage of the chosen method is the rigidly structured process with predefined answers, which does not allow for individual follow-up questions to be asked or issues to be explained if necessary [Hie09, p. 116]. For scientific compliance, each participant had to agree to a scientific evaluation of the information provided (see Appendix C). The participants were also always able to cancel the survey and were not forced to give answers in any way.

Questions The hypotheses were derived from the process defined in Section 6.1, and the questions were generated by deduction and tested with the help of empirical research. Complications within each activity of the abstract model (see Figure B.4) were assumed. The detailed models helped to derive possible answers (see Appendix B). From the perspective of empirical research, this is an ex post facto design because both independent and dependent variables are measured. The questionnaire was programmed so that questions that did not match the criteria were not shown to the respondent. For example, a user who has not filed a request for erasure was not asked how the performance o the process was. Also, the basic questions lead to a drop-out in case the respondent did not match the population criteria (see next section).

Population The possible population of the survey at hand is every citizen of the European Union. The population of the survey were all people able to file GDPR requests. Random sampling was used to ensure representativeness. In order to limit any local influences and to conduct the research in a meaningful way, a restriction was made on people from Austria. The research was a cross-sectional study in which different people in different groups (e.g. age or gender) were examined at the same time. Furthermore, this empirical study was (currently) only conducted once; a time-based hypothesis or trend could, therefore, not be analysed. Also, according to the hypothesis, a restriction was made on people who had already gone through this process. Accordingly, in addition to the basic data, there were a number of drop-out criteria in the resulting questionnaire that led to the survey being abandoned. These were:

1. Did not agree with a scientific evaluation of the information provided.
2. Was not Austrian
3. Has never carried out one of the evaluated processes

The sample was chosen randomly selection of individuals from the population. After the drop-out criteria, n=99 agreed to the terms and finished the questionnaire completely. Basic data were collected to verify the validity of the sample.

Data Collection For data collection of the quantitative method, an online questionnaire was used. Through the standardised questionnaire, all respondents receive the same fixed questions and answer options. This enabled a uniform interview situation for the entire sample [Sti05, p. 135]. Depending on the question, either a single or multiple selections were possible. A Likert scale was used for questions of grading. The sorting of the questions was randomised for lists (i.e. not Likert scale questions) in order to prevent a bias. For the survey creation and answering, the online platform SurveyMonkey [2] was used because of its ease of use and GDPR compliance [Sur21].

The measurement of the facts of the respective term has to be defined and aligned according to the respective scale level [May09, p. 72]. The theoretical terms were

[2] https://www.surveymonkey.com

analysed in different dimensions in order to check whether essential aspects are included, which later serve the question creation [May09, p. 34].

Pre-Test The questionnaire was subjected to a pre-test [Hie09, p. 116] before it was put online. In this way, errors such as misunderstanding a question or ambiguities were counteracted in advance.

6.2.2 Analysis

The resulting data were exported and analysed. This allowed applying various statistical methods for the evaluation and analysis of the data. Through the descriptive analysis methods, the distributions and correlations within the sample were presented [May09, p. 113]. The selection or application of the respective procedures depended not only on content considerations but also on the measurement or scale level of the respective variables. The level is derived from the scale types "nominal, ordinal, interval, and ratio scale" [Sti05, p. 247].

Methods of descriptive statistics, frequencies in per cent and number, means, standard deviations, as well as minimum and maximum and weighted average, were used to describe the data. Histograms, bar charts and occasionally pie charts were used for graphical representation. The latter was not included in the thesis.

The demographic composition of the respondents (after exclusion based on the exclusion criteria) can be summarised as follows:

- The respondents were 71,7% male and 28,3% female (see Figure C.5) (see Figure C.1).
- Most respondents (82,0%) were between 18 and 39 years old (see Figure C.3).
- Most respondents (72,8%) had an Advanced technical college certificate or high school diploma (Matura) or higher (see Figure C.4).

Regarding the execution of their rights, 16,2% have already requested an erasure (Art. 17 GDPR) multiple times and 64,7% once. However, only 64,6% made a request for information (Art. 15 GDPR) once and 8,1% several times. This shows that the right to be forgotten (Art. 17 GDPR) has a repetition rate almost twice as high as a request for information (Art. 15 GDPR), even though the solo request rate is nearly equal. The repetitive users can be seen as 'experienced' within the process, and therefore their answers might be more expressive.

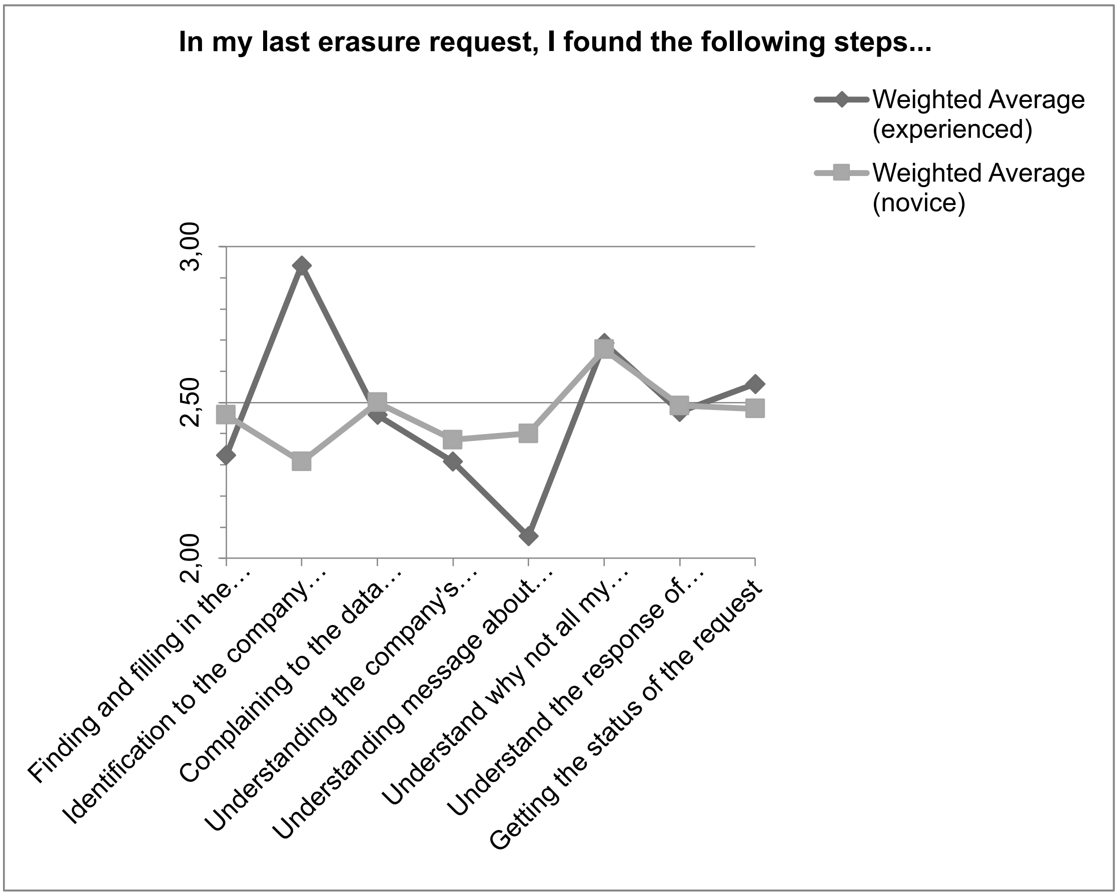


Figure 6.4: Comparison of difficulty of steps in the request for the erasure of experienced and novice participants

A comparison of the answers given on the difficulty of steps between ‘experienced’ participants (that did a request multiple times) and ‘novice’ participants (that made only one request) shows that the anticipated difficulty in the steps ‘Identification to the company (e.g. scan ID card)’ and ‘Understanding message about delay’ were significantly different (see Figure 6.4). ‘Experienced’ participants find the method of identification more difficult and the messages about delays easier.

When drilling further into the data of ‘experienced’ participants in those two questions, the specifics of the issues are apparent. The ‘Identification to the company (e.g. scan ID card)’ was most difficult (3,7) when the request was submitted ‘In an online form of the company’ or ‘Informally by email or mail’ (2,6). The others were rated with 2,0 (see Table 6.1). This leads to the conclusion that identification is more difficult when using the company’s platform or informal communication, such as email.

Table 6.1: Answers to how difficult identification was depending on the way of filing the request in experienced participants

	very simply	simply	ok	difficult	very difficult	was not the case	Total	Weighted Average
-	0,00%	14,30%	28,60%	28,60%	28,60%	0,00%	43,80%	3,7
Q9: In an online form of the company.	0	1	2	2	2	0	7	
-	0,00%	60,00%	20,00%	20,00%	0,00%	0,00%	31,30%	2,6
Q9: Informally by e-mail or post	0	3	1	1	0	0	5	
-	33,30%	33,30%	33,30%	0,00%	0,00%	0,00%	18,80%	2,0
Q9: By means of a phone call	1	1	1	0	0	0	3	
-	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	6,30%	2,0
Q9: By means of a form by e-mail or post	0	1	0	0	0	0	1	

The same two ways of requests had other issues: receiving the status, Locating and completing the application form.

The 'Complaint to the data protection authority' was with both 'novice' and 'experienced' participants most difficult when it was filed 'By means of a form sent by e-mail or post' (see Table 6.2).

Table 6.2: Difficulty of complaint to the data protection authority for both novice and experienced participants

	very simply	simply	ok	difficult	very difficult	was not the case	Total	Weighted Average
-	11,40%	37,10%	22,90%	8,60%	8,60%	11,40%	43,80%	2,61

Q9: In an online form of the company.	4	13	8	3	3	4	35	
-	16,10%	32,30%	32,30%	6,50%	0,00%	12,90%	38,80%	2,33
Q9: Informally by e-mail or post	5	10	10	2	0	4	31	
-	20,00%	30,00%	30,00%	10,00%	0,00%	10,00%	12,50%	2,33
Q9: By means of a phone call	2	3	3	1	0	1	10	
-	0,00%	0,00%	50,00%	25,00%	0,00%	25,00%	5,00%	3,33
Q9: By means of a form by e-mail or post	0	0	2	1	0	1	4	

When drilling into the issue of ‘Understanding message about delay’ by ‘novice’ users found in Figure 6.4, the reason might lie in education. Figure 6.5 shows the education of ‘novice’ (Q7 answered with ‘once’) against the education of ‘experienced’ (Q7 answered with ‘several times’). This leads to the thesis that education might be the hidden reason for the difficulty in understanding messages about delays.

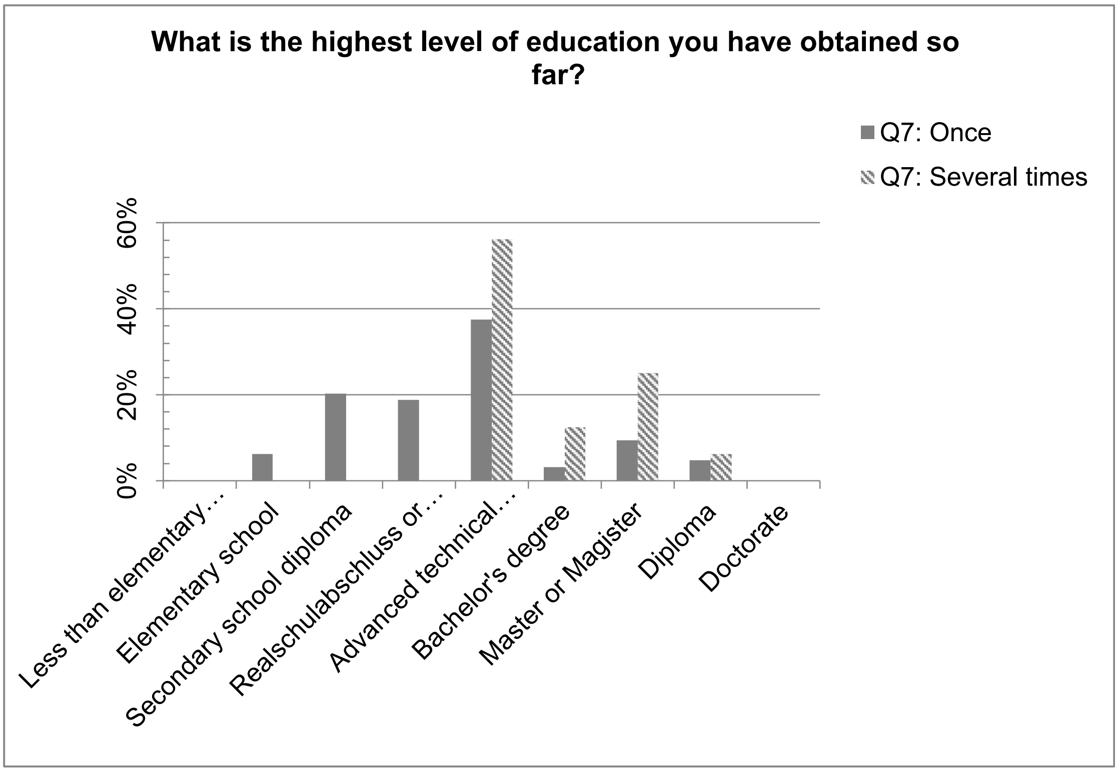


Figure 6.5: Comparison of education of experienced and novice participants

All participants that have requested information (Art. 15 GDPR) have made at least one request for erasure. During the requests for information 'novice' participants assessed all steps more difficult than 'experienced' participants but one; namely the complaint at the DPA (see Figure 6.6).

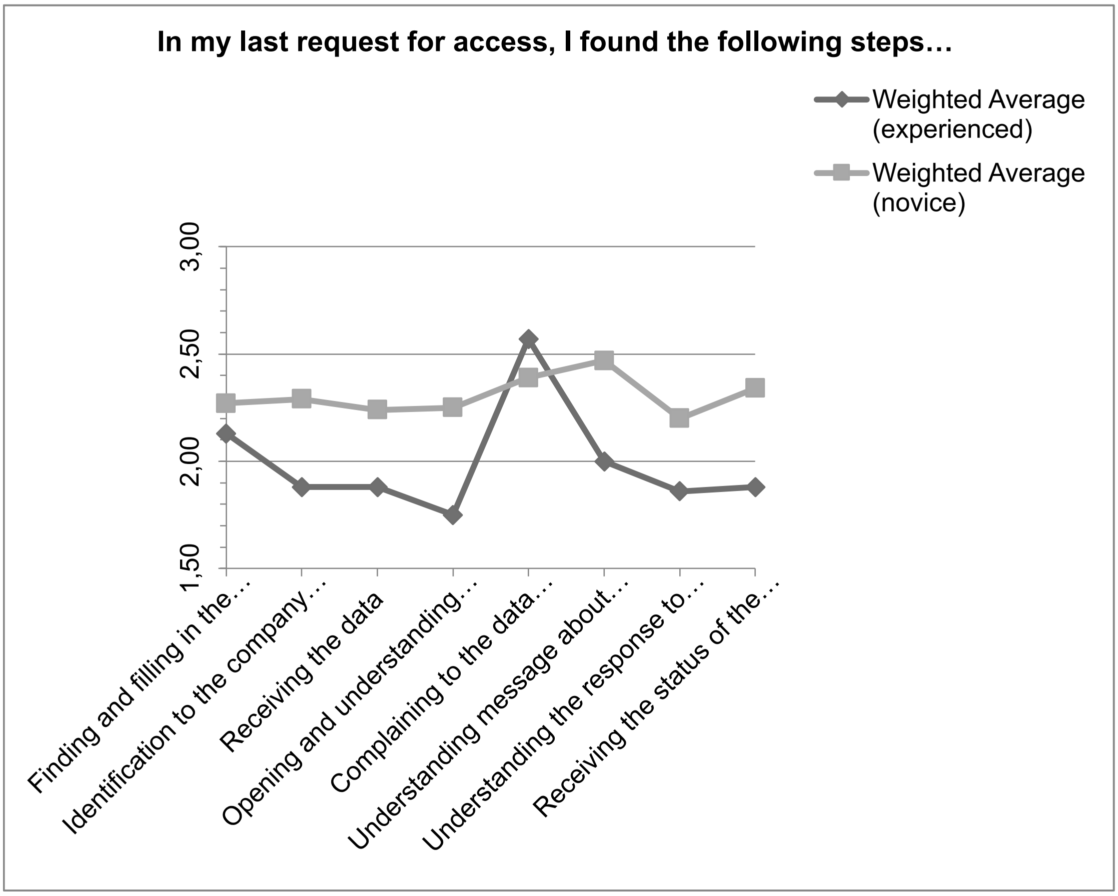


Figure 6.6: Comparison of difficulty of steps in the request for information of experienced and novice participants

When taking all experience levels into account, the most difficult steps were ‘Complaining to the data protection authority’, ‘Understanding message about delay’ and ‘Receiving the status of the request’.

The usual wait time for a request for information (Art. 15 GDPR) and request for erasure (Art. 17 GDPR) is between one day and four weeks (see Figure 6.7 and 6.8). Still, 12 (15,0%) waited more than four weeks for an answer to their request for erasure, and 2 (2,5%) did not receive an answer at all.

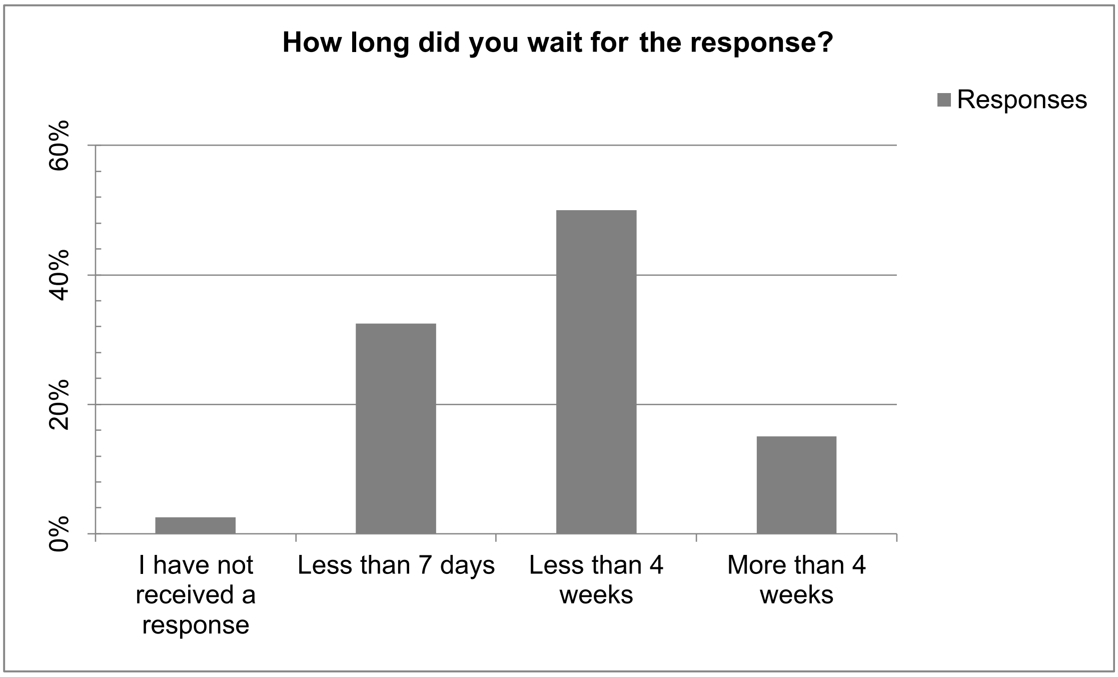


Figure 6.7: Time participants had to wait for their answer to their request for erasure

9 participants (12,5%) waited for their request for information for more than four weeks for an answer, 1 (1,4%) did not receive an answer at all (see Figure 6.8).

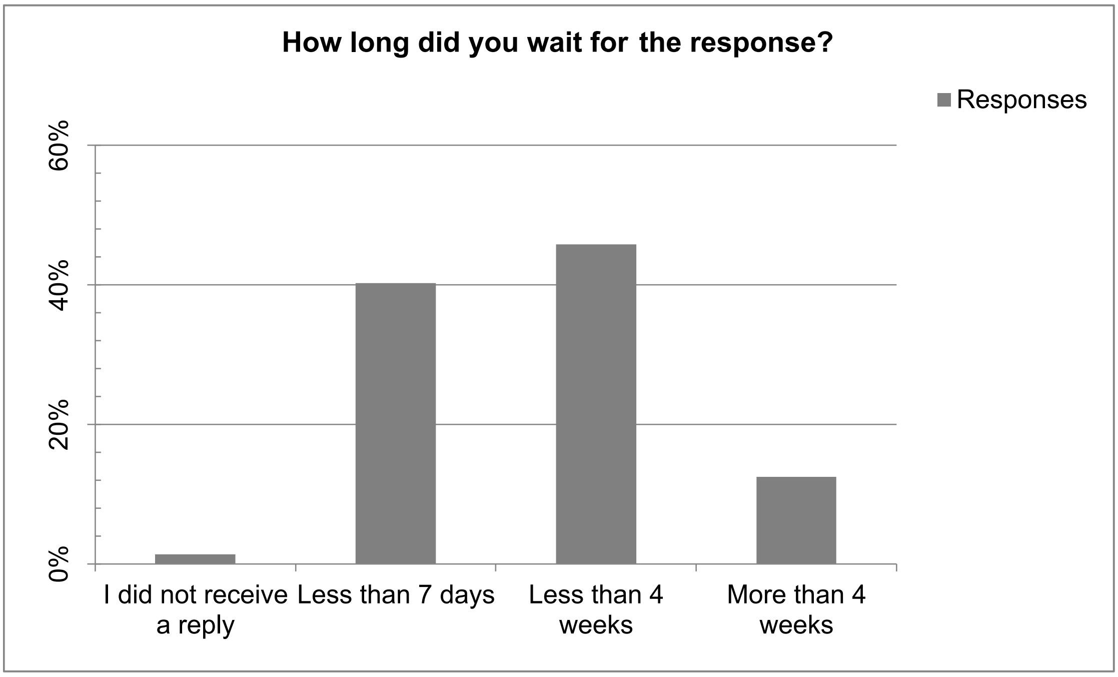


Figure 6.8: Time participants had to wait for their answer to their request for information

The response of requests was mostly (58,8%) received via email (see Figure 6.9).

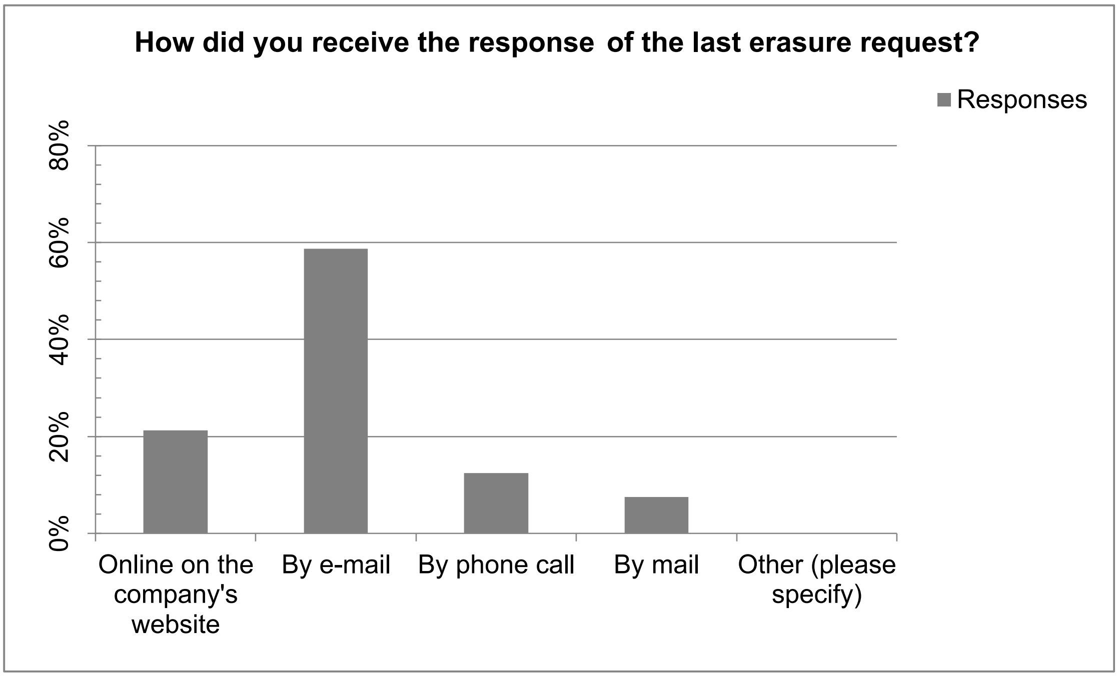


Figure 6.9: Response channel of the request for erasure

The response to a request for information was mostly (52,8%) answered with a structured data file (e.g. Excel, CSV or similar).

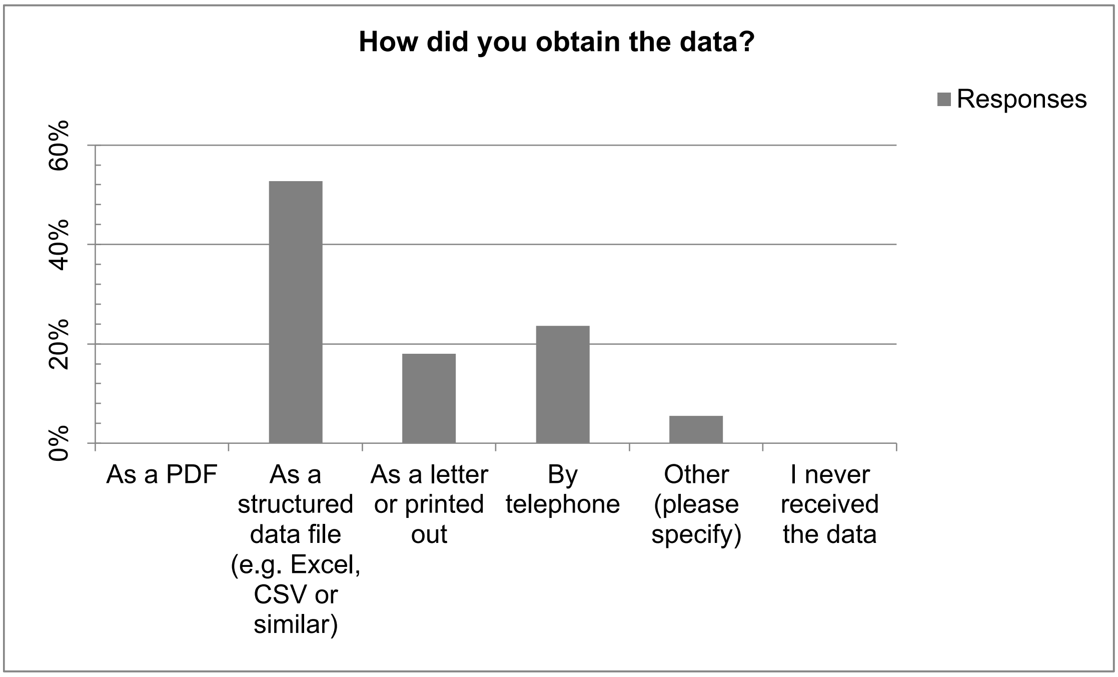


Figure 6.10: Response type of the request for information

Finally, the participants were asked what their most important factors were when requesting information. The question was a ranking question. The participants needed to sort the possible answers accordingly. The score for each answer choice was calculated to analyse which answer choice was most preferred overall. The answer choice with the largest score is the most preferred choice. The score is calculated as follows: n is the number of positions. w is the weight of ranked position i. The higher a position it is ranked, the better $(1 + n - i)$. x is the number of responses for the answer choice in the ranked position i.

$$f(\vec{w}, \vec{x}) = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} x_i} \tag{6.1}$$

	Score	**Minimum value**	**Maximum value**	**Median**	**Mean**	**Standard deviation**
that the application is understandable	5,6	1,0	9,0	4,0	4,4	2,5
that my data is secure	5,4	1,0	9,0	5,0	4,6	2,7
that I have the possibility to complain	5,3	1,0	9,0	5,0	4,7	2,5

that I have personal contact	4,9	1,0	9,0	6,0	5,1	2,6
that I can see the status of the processing	4,8	1,0	9,0	5,0	5,2	2,4
that I can do it from my smartphone	4,8	1,0	9,0	5,0	5,2	2,8
that I can also submit it in paper form	4,8	1,0	9,0	5,0	5,2	2,5
that it can be informal	4,8	1,0	9,0	5,0	5,2	2,7
that little data is requested	4,7	1,0	9,0	5,0	5,3	2,4

Table 6.3 shows the by score sorted results of the factors participants rated the most important. The three most important factors were that the application is understandable, that my data is secure, and that I have the possibility to complain.

These factors need to be considered when designing solutions for data protection, according to this study.

6.2.3 Conclusions

A survey with 99 participants, of which 16 were experienced in the means of having already made multiple requests for erasure (Art. 17 GDPR), resulted in the identification of the following issues:

- Understanding message about the delay (by uneducated)
- Identification when using mail and platforms
- Complaint to the data protection authority

The same survey showed the requests for information (Art. 15 GDPR) as the following issues:

- Complaining to the data protection authority
- Understanding message about the delay
- Receiving the status of the request

Also, anticipated factors of a well-implemented process for information requests were ranked with the following top three factors:

1. that the application is understandable;
2. that my data is secure;
3. that I have the possibility to complain about.

These issues and factors for good solutions shall help to define a qualitative solution space.

6.3 Qualitative Definition of the Solution Criteria

It must be elaborated on which requirements are to be fulfilled for the activities that showed certain issues. This must then lead to a measurable catalogue of criteria questions to check the degree of fulfilment of these requirements.

6.3.1 Design

Several qualitative methods would have been appropriate to be used (see Section 5.3) to investigate these requirements.

The Delphi technique is a problem-solving tool in which a panel of experts provides feedback and ideas on a specific topic or issue [MVHA07]. The main feature of the Delphi Technique is that it minimises creativity-reducing mutual influence and peer pressure, but still allows for collaboration, and a common consensus on the solutions exists at the end. It takes its name from the Oracle of Delphi, a priestess in ancient Greece who was known for her ability to make accurate prophecies. The comparison is taken because, in contrast to ex-post surveys, i.e. those in which an event that has already happened is surveyed, the future is assessed.

The 'classical' Delphi technique begins with a questionnaire or set of problems presented to a group of experts [KMH10]. The experts then anonymously submit their ideas and solutions [KMH10, MVHA07]. These submissions are then filtered and summarized [KMH10, MVHA07]. In the second round, they are again anonymously given to the experts in the form of a questionnaire [KMH10]. The experts then decide on the best solution or sort the solutions by quality [KMH10]. The results are compiled again in subsequent rounds and shared anonymously with the group until a threshold or consensus is reached [KMH10].

The Delphi technique was used within this process to find a consensus between the experts on what criteria are used to evaluate the success or fulfilment of requirements of a supporting technology (solution criteria) for a certain activity.

The concrete implementation of the rounds was designed as follows:

- Round 1: Qualitative questioning about criteria for an activity (free response as text). Open questions (no fixed answer options). Hardly any operationalisation (general question). Aim: Collection of ideas.
- Preparation 2nd round: Extraction of criteria and summary of similar criteria.
- 2nd round: Qualitative and quantitative questioning of validity and importance. Sorting of criteria for an activity (fixed answer options). Space for argumentation of sorting.
- Preparation of third round: Re-sorting of criteria according to answers. Summary of the arguments.
- Round 3: Qualitative and quantitative questioning according to importance. Sorting of criteria for an activity (fixed answer options). Space for argumentation of the sorting.

The rounds are repeated until a standard deviation below 3.0 is reached (termination criterion).

Population In the course of this qualitative research, the 'Theoretical Sampling' [GS17] was used for a sample selection. For this purpose, ten domain experts (n=10) in the field are selected at the beginning of the research process. The experts shall be proficient in implementing solutions and defining criteria for solutions.

Since both data protection and product development aspects are to be covered, experts in both fields will be interviewed. The criteria are formally defined by:

- Has at least four years of proficiency in the legal aspects of data protection and basic knowledge of requirements engineering and product development
- Or has at least four years of proficiency in requirements engineering and product development and basic knowledge of legal aspects of data protection

The experts received pseudonyms on each round of the questionnaire to prevent de-anonymisation. The pseudonyms (expert numbers) are used within all documentation. All additional possible privacy-influencing details have been censored in the documentation. With this precaution, the privacy of the individuals shall be protected. Also, the experts are from very different groups (not employed at the same company, institute or likewise) and were not introduced to each other or made aware of who was participating.

Data Collection The qualitative and quantitative rounds have been done with an online questionnaire. The first round was done with simple input fields (free-text fields). The subsequent rounds were done with a sort field for the rating and a free-text field for the comments. This enabled a uniform interview situation for the entire sample [Sti05]. For the survey creation and answering, the online platform SurveyMonkey [3] was used because of its ease of use and GDPR compliance [Sur21].

Questions The questions were uniformly designed according to the business process defined in Section 5.1. For the questions, the following question template was used "What criteria would you use to evaluate the implementation of [Activity] in the course of a [Process Name]?". The full questionnaires can be found in Appendix D.

After the first round, the questions were designed according to the answers given (see Appendix D). The first round specifically asked for factors; the following rounds asked to rank the factors - until a consensus was found.

The questions in the first round were open questions. Only the two GDPR processes were shown and explained, and criteria for software supporting the activities were asked.

The first test round was done with a test subject that checked whether the questions were clear and tried to answer the first two questions, and found the following errors (sorted by severity):

[3] https://www.surveymonkey.com

Table 6.4: Thematic analysis results for the activity ‘Create Request for Deletion’

Thematic map leaves for S_1_1 `Create Request for Erasure’		
S_1_1	Usability	User friendly selection which data shall be covered by the request for erasure
S_1_1	Speed	Time it takes to discover the information.
S_1_1	Accessibility	Accessibility of communication also for disabled users.
S_1_1	Confirmation	Timely generated confirmation message.

- The activities were not clear enough described
- The process image was too small to read (on certain devices)
- One phrasing was unclear in the guidance

The found issues were fixed, and the questions were sent to the expert participants. They were given a time limit of two weeks which was extended by another week.

The answers were coded and clustered with the help of mindmaps [BC06]. First, the raw answers (see Appendix D) were visualised and split up in atomic reasoning. Then they were sorted by similarity, and the overarching term (‘Criterion’) was searched. This was done similarly to the thematic analysis by Braun and Clarke [BC06]. The results of the interviews were accordingly developed, and the final thematic map was created with the help of the aforementioned criteria. A common visualisation method for thematic maps is a mind map. In the case of analysis of activities regarding the themes, a branch was created for each activity. Therefore, the findings regarding the activity were either a leaf in the mind map or branches as sub-categories with leaves.

An Excerpt of the results for the activity ‘Create Request for Deletion’ is shown in 6.4.

Synonyms were filtered, and criteria were merged if they were the same (see Appendix D). Then questions were created asking for the quality of the criterion. These criteria questions were again sent to the participants to sort them by importance for each activity (see Appendix D). The results are then summarised in tables and meta-data of the business process diagram.

6.3.2 Analysis

After the criteria had been gathered, the criteria questions were created and sorted by the experts. The ranking of each individual expert was summed up, and a ranking score (see Equation 6.1) was calculated.

This score represents how important the criterion and the responding criteria question are.

Table 6.5: Criteria questions for activity 'File Request for Erasure'

Activity S_1_1 'Create Request for Erasure'		
ID	**Rank**	**Question**
S_1_1.Q1	5,7	How easy can the deletion form be found?
S_1_1.Q2	5,1	How much time is needed to file the request?
S_1_1.Q3	3,7	How well can the data to be deleted be defined?
S_1_1.Q4	3,2	How well is the process for creating request for erasure explained?
S_1_1.Q5	1,8	How well is the use of templates when creating a request?
S_1_1.Q6	1,5	How well does the form check on completeness of information when creating a request?

After reaching a consensus, the results were added to the activity information in the process. Table 6.5 shows the result of the first activity, 'Create Request for Deletion', which represents the criteria questions as an example of the results.

The resulting criteria question catalogue was then formulated and reviewed again. The criteria catalogue was then used for the expert interviews.

The application was introduced to the experts. Then they were asked about the application properties by answering the questions given to them.

Finally, the results were visually presented. The final artefacts were then:

- A BPMN Diagram of the processes of the regulation
- Criteria Questions connected to the activities
- Results of the analysis regarding the application and evaluation

6.3.3 Conclusions

The BPMN Diagram of the processes of the regulation can be used to understand the general process in the regulation, especially for development. It can also be used by people who want to understand the legal options and as a reference guide on how to exercise their rights.

The criterion questions are granular and deep in detail. Therefore, they are relevant for people concerned with concrete tasks in very specific sub-areas. They are not suitable for a quick overview. People who deal with criterion questions will typically be subject matter experts.

As a final result, the annotated BPMN was used to further improve the application under evaluation. It showed the quality of the supportive properties and, therefore, the potential for improvements and extensions.

6.4 Presentation of Case Studies

The following case studies were chosen to be assessed with the presented process. The case studies are used as an empirical research method that is intended to investigate the proposed assessment process in its real environment. Each case represents software to support parts of the presented GDPR processes.

These case studies will be presented based on the problem to be solved and their solution description. After the presentations of the case studies, they will be evaluated in terms of their impact on GDPR process activities.

6.4.1 Case Study: Citizen Empowerment Platform for Requests for Erasure using Blockchain Technology

The General Data Protection Regulation (GDPR) grants European citizens the right to be forgotten (see Chapter 2). A citizen can request information from a controller whether their personal data is processed and eventually request to erase it from the controller and its processors. Pursuant to [Cou16a, Art. 17], the data subject has the right to demand that the controller erase the personal data concerning him or her.

The controller is then obliged to either fulfil the request without delay [Cou16a, Art. 17 (1)] or explain why it will not fulfil the request. This can be for various reasons. A few of them are: that it cannot clearly identify the data subject, that it has already anonymised the personal data, that it needs the remaining personal data for other legal obligations or that the data subject has requested too often.

Either way, the controller must answer the data subject within one month [Cou16a, Art. 12 (3)]. This answer can also be a message that the erasure might take longer (see 2.3.5). In case the request takes too long, the data subject has the right to file a complaint with their local data protection authority. Ideally, this is not necessary, and the controller and its processors erase the personal data or anonymise it and respond to the request with a success message. The maximum time to be taken is defined as three months after receiving the request [Cou16a, Art. 12 (3)].

Problem Statement of the Case Study

The problem to be solved is requesting the erasure from the controller and proofing a request in case of a complaint with the data protection authority. Several issues could occur during this process. The first problem arises when the data subject must find the right contact or communication channel to contact the controller. This can be difficult to achieve since direct contacts are oftentimes hidden on web pages.

The second problem arises when the data subject needs to know the contents of such a request to comply with legal requirements.

The data subject might be confronted with the requirement to send identification to the data controller in order to verify the identity. The problem arising from this request is that it usually is not secure to send a scan of a legal ID to a third party.

Another possible problem might occur when the data subject needs some proof that he or she filed a request in order to file a solid complaint at the DPA (even though it is not legally required).

Ultimately, it is complicated to file a complaint with the data protection authority in terms of the necessary information on the request for erasure and other information needed.

Description of the Solution

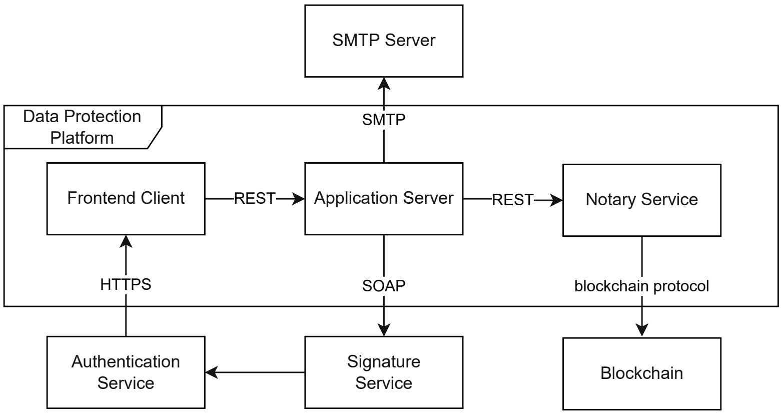


Figure 6.11: Architectural overview of the system [SPB+20]

The author's research proposed a solution to these problems with the help of an application using blockchain technology [SPB+20]. The system would help data subjects quickly find companies' data protection contacts and enable them to file a digitally signed erasure request. It furthermore allows proof of the request to a DPA in case of a complaint with the help of a public Blockchain. It also reminds the data subject to create a complaint in case the request takes too long.

The stated architecture allows tracking of the request without needing to place personal data on the blockchain. It uses an electronic IDentification, Authentication and trust Services (eIDAS) compatible signature mechanism.

Figure 6.11 shows the architecture of the proposed system. It consists of a classic web application with a front end using web technologies and a backend (application server). The backend offers the frontend a REST interface with the help of which all operations can be carried out. The backend uses a signature service to sign PDF files. This signature service is connected to an authentication service (from A-Trust) in an overall system. This overall system can be accessed via SOAP and offers the possibility to have a PDF authenticated and signed by the user's qualified signature via HTTPS forwarding.

A notary service (OpenTimeStamps), which in turn uses the Bitcoin blockchain, is used for the persistent, unchangeable storage of document fingerprints and their creation time. It timestamps the hash value, puts it in a Merkle hash tree, and stores the result on the Bitcoin blockchain. Therefore, no personal data is stored on the blockchain, and the existence of the request can be proven later. The Merkle tree is a hierarchic hash of data (see Section 3.3) that is used to reduce transaction costs. With the help of the Merkle tree, only one hash must be persisted on the blockchain in order to prove the existence of many documents at a time. The backend also communicates with an SMTP server. It uses the SMTP server to send emails and to access the communication protocol in order to store them as further proof of the request.

User Interaction to file a Request for Erasure The application for filing a request for erasure with a controller is a website available on the secure world wide web. To access the website, no login or registration is needed. The request starts with an explanation of the user's rights (the user is called 'data subject' in this context) and the process of filing an erasure request.

The request is created with a wizard, meaning the user is guided through the creation process. The steps of the wizard are depicted on the top of the web page. Within the request form question, mark symbols show further information regarding it.

Initially, the data subject can choose a company or enter their contact information manually. The fields of the form match the corresponding PDF form of the Austrian Data Protection Authority (DPA).

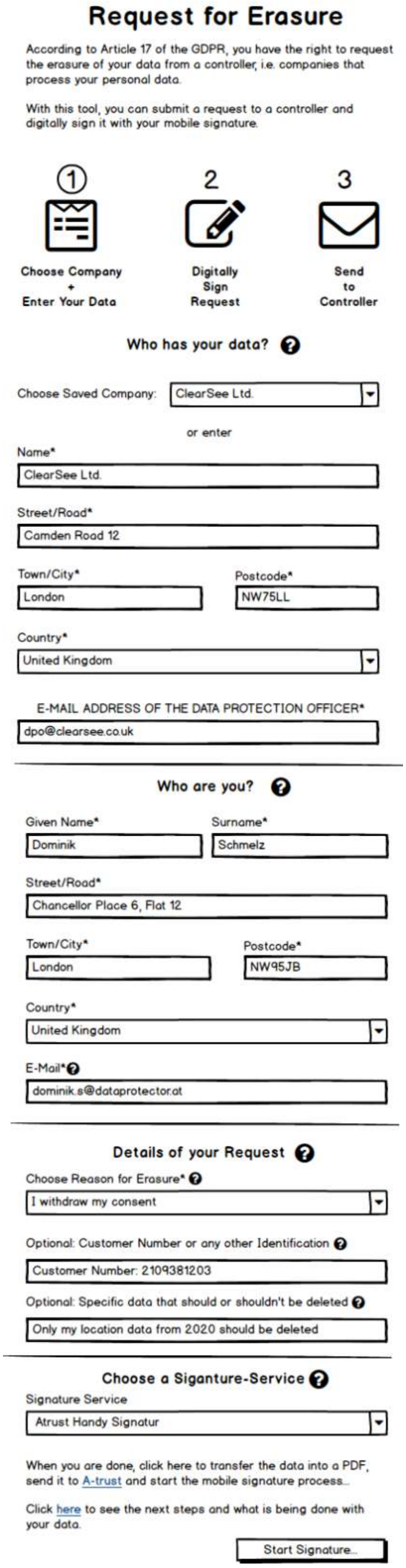
The contact information is manually managed by an administrator.

The data subject then enters its contact information and needs to specify a reason for the erasure request. The requirement for this information has been derived from the required fields in the request form of the Austrian data protection authority.

The data subject can specify an identity known to the controller, such as a user identification or customer number or any other means of recognition for the controller, to allow the controller to select the data of the data subject within its database.

Before submitting the form, the data subject receives an explanation of what happens with the provided data and the next steps.

Technical detail: The information provided by the data subject is put in the PDF mentioned above form of the Austrian data protection authority and transmitted to the A-trust servers as a signature request. The A-trust server returns a URL that must be given to the data subject to authorise the qualified signature.



Request for Erasure

According to Article 17 of the GDPR, you have the right to request the erasure of your data from a controller, i.e. companies that process your personal data.

With this tool, you can submit a request to a controller and digitally sign it with your mobile signature.

Who has your data?

Choose Saved Company: ClearSee Ltd.

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER*: dpo@clearsee.co.uk

Who are you?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom

E-Mail*: dominik.s@dataprotector.at

Details of your Request

Choose Reason for Erasure*: I withdraw my consent

Optional: Customer Number or any other Identification: Customer Number: 2109381203

Optional: Specific data that should or shouldn't be deleted: Only my location data from 2020 should be deleted

Choose a Siganture-Service

Signature Service: Atrust Handy Signatur

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Click here to see the next steps and what is being done with your data.

Figure 6.12: Request for Erasure: Step 1 - Enter Your Data

After submitting the information, the data subject must authenticate with the chosen (eIDAS) authentication service. In this case, its Handy-Signatur (National mobile identification of Austria). After the two-factor authentication, the data subject authorises the qualified signature of the PDF.

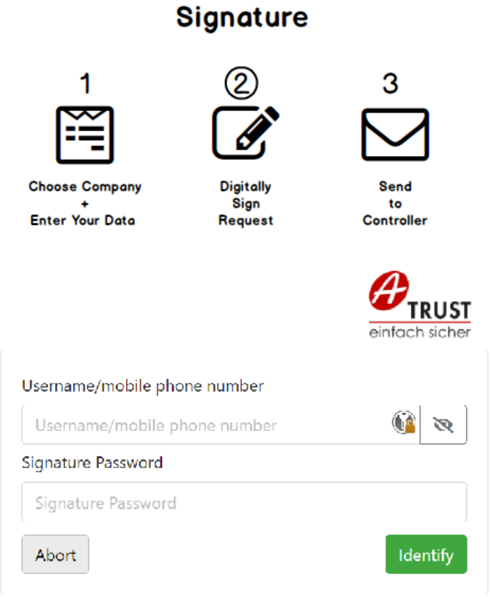
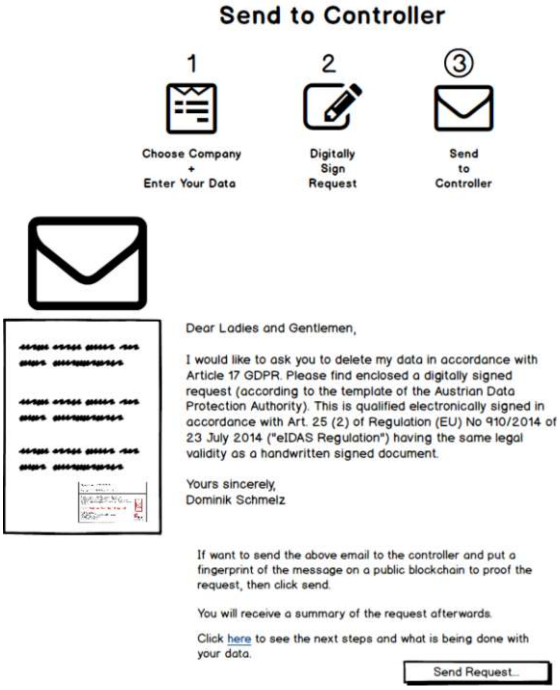


Figure 6.13: Request for Erasure: Step 2 - Signature

A summary is shown with the signed PDF, and the email sent to the controller. If the data subject agrees to send the email and the PDF attached, the mail is sent to the controller, and a fingerprint is stored on a public blockchain.

The email is sent to the system's own SMTP Server, and the request (including the mail and PDF) is hashed and put on a public blockchain. The free OpenTimestamps service is used for this service. It timestamps the hash value, puts it in a Merkle hash tree, and stores the result on the Bitcoin blockchain. Therefore, no personal data is stored on the blockchain, and the existence of the request can be proven later.



Send to Controller

Dear Ladies and Gentlemen,

I would like to ask you to delete my data in accordance with Article 17 GDPR. Please find enclosed a digitally signed request (according to the template of the Austrian Data Protection Authority). This is qualified electronically signed in accordance with Art. 25 (2) of Regulation (EU) No 910/2014 of 23 July 2014 ("eIDAS Regulation") having the same legal validity as a handwritten signed document.

Yours sincerely,
Dominik Schmelz

If want to send the above email to the controller and put a fingerprint of the message on a public blockchain to proof the request, then click send.

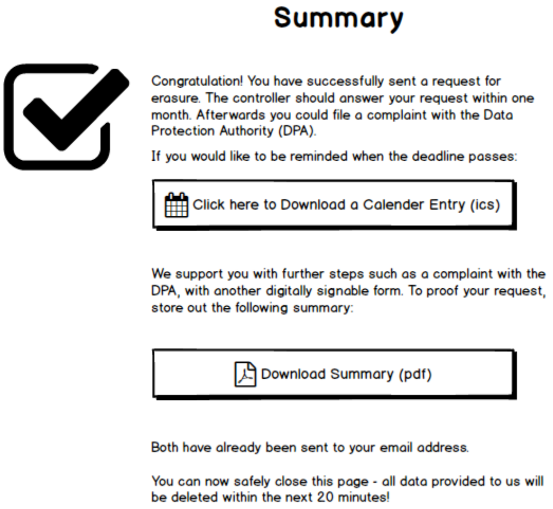
You will receive a summary of the request afterwards.

Click here to see the next steps and what is being done with your data.

Figure 6.14: Request for Erasure: Step 3 - Send

The data subject receives a success message on the web page and confirmation via email. The message and email include a summary of the request as a PDF and a calendar entry (ICS file) to be stored in the data subjects calendar software. The calendar entry reminds the data subject of the date it can file a complaint with the data protection authority if the controller has not answered the request. In case of a delay, the data subject must receive a notification from the controller; in this case, the data subject must move the calendar entry manually.

The calendar entry includes a link to a form to fill out the complaint information and send it again digitally signed to the Data Protection Authority.



Summary

Congratulation! You have successfully sent a request for erasure. The controller should answer your request within one month. Afterwards you could file a complaint with the Data Protection Authority (DPA).

If you would like to be reminded when the deadline passes:

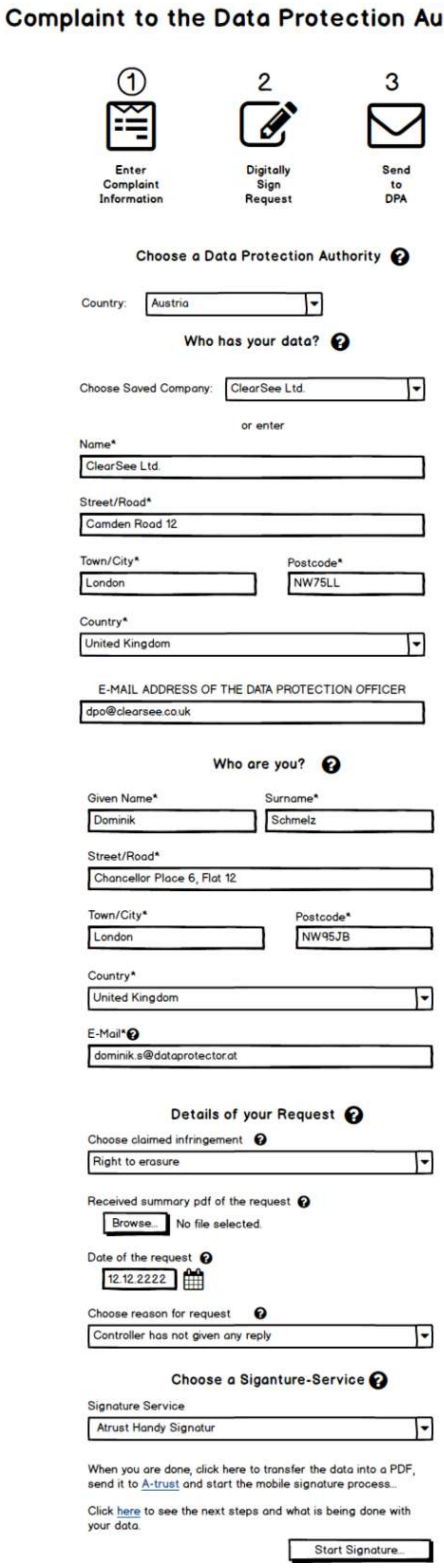
We support you with further steps such as a complaint with the DPA, with another digitally signable form. To proof your request, store out the following summary:

Both have already been sent to your email address.

You can now safely close this page - all data provided to us will be deleted within the next 20 minutes!

Figure 6.15: Request for Erasure: Step 4 - Summary

To file a complaint, the data subject can click the link provided in the calendar entry. The data subject can choose a data protection authority it wants to address. It must specify the details of the request. These have been derived from the Austrian data protection Authorities complaint form. The summary received via email can be attached to the complaint to prove the request to the authority. Many fields of the complaint form are automatically filled out (such as type of signature, type of contact etc.).



Complaint to the Data Protection Authority

Choose a Data Protection Authority ?

Country: Austria ▾

Who has your data? ?

Choose Saved Company: ClearSee Ltd. ▾

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom ▾

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER: dpo@clearsee.co.uk

Who are you? ?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom ▾

E-Mail* ?: dominik.s@dataprotector.at

Details of your Request ?

Choose claimed infringement ?: Right to erasure ▾

Received summary pdf of the request ?: Browse... No file selected.

Date of the request ?: 12.12.2222

Choose reason for request ?: Controller has not given any reply ▾

Choose a Siganture-Service ?

Signature Service: Atrust Handy Signatur ▾

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Click here to see the next steps and what is being done with your data.

Start Signature...

Figure 6.16: Request for Erasure: Optional Step 5 - Complaint

6.4.2 Case Study: Securing GDPR Data Transfers with Blockchain Technology

The General Data Protection Regulation (GDPR) protects the personal data of citizens of the European Union. It gives people the right to know what personal data is being collected about them and to access the data being processed [Cou16a, Art. 15] including purposes of processing activities; categories, source, recipients, retention time of personal data; and information of rights and whether automated decision-making or profiling is done. [Cou16a, Art. 15 (3)] also grants the right to request a copy of the data processed.

The right to data portability [Cou16a, Art. 20] also allows the data subject to request its personal data. It covers personal data the data subject has provided to the controller [Cou16a, Art. 20]. The data subject has the right to receive these data in a structured, common, machine-readable format [Cou16a, Art. 20 (1)] and has the right to ask for a transfer to another controller were feasible [Cou16a, Art. 20 (2)].

Problem Statement of the Case Study

Within the request for access, several problems might arise. The data subject might not have the technical knowledge to file the request or to find the means of contact with the data controller. The specific problem with these kinds of GDPR requests is the transportation of the data. The data controller or processor needs to transfer the stored data to the data subject in a secure and data-protected way since the data contains personal data.

The data subject might need to identify itself with the data controller and file a request to access certain information. Again the data subject might want to file a complaint at the DPA if the request takes too long or is not handled correctly.

Description of the Solution

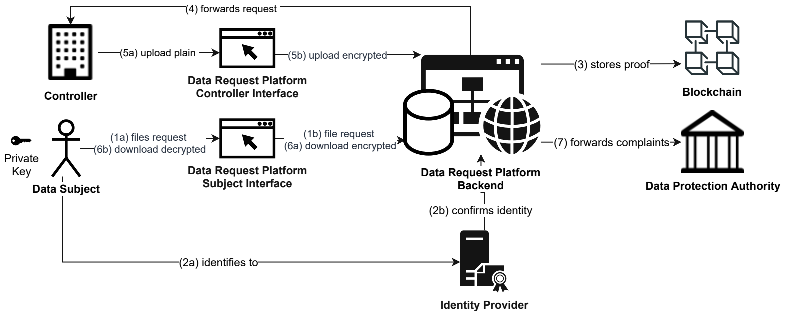


Figure 6.17: Architectural overview of the system [SPNG21]

In order to enforce these rights, the author [SPNG21] has proposed a system in which people can submit requests to data controllers (i.e. the organisations that collect and use personal data) in a standardised way. This system would use blockchain technology to create a record of the request that could not be changed or deleted.

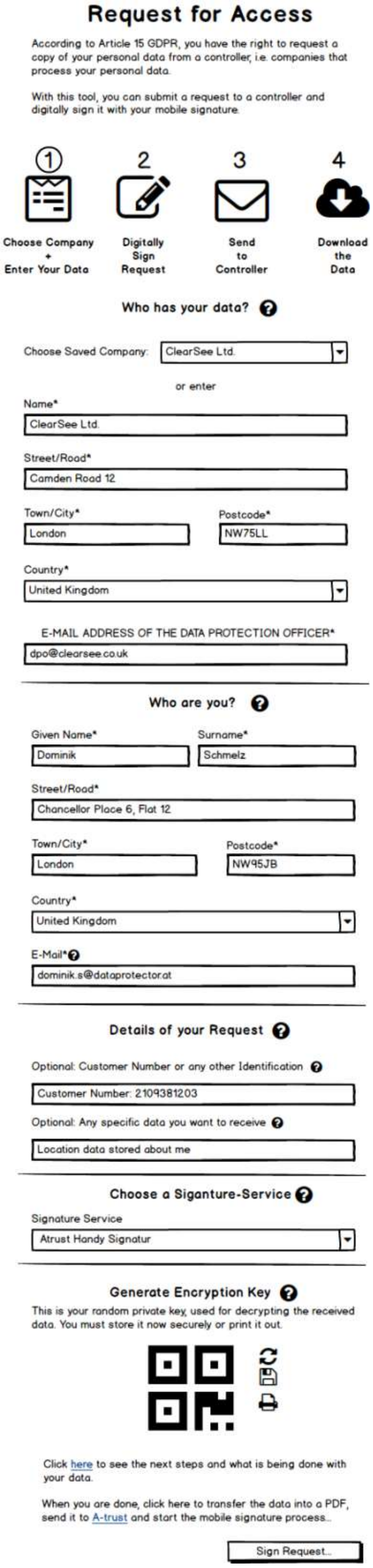
The system would also help data controllers to comply with the GDPR by providing a way for them to easily respond to requests from data subjects.

The authors believe that this system could be implemented on a multinational level in order to help protect the data of people all over the world.

User Interaction to file a Request for Access The application for filing a request for access with a controller is a website available on the open Internet (HTTPS). To access the website, no login or registration is needed. The request starts with an explanation of the user's rights (the user is called 'data subject' in this context) and the process of filing a request for access.

The request is created with a wizard, meaning the user is guided through the creation process. The steps of the wizard are depicted on the top. The request form shows further information regarding the topic as help icons (question marks). Initially, the data subject can choose a company or enter their contact information manually. The fields of the form are derived from the corresponding form of the Austrian Data Protection Authority (DPA). The contact information is manually managed by an administrator.

The data subject can specify an identity known to the controller, such as a user identification or customer number or any other means of recognition for the controller to allow it to select the data of the data subject within its database. The data subject's web browser automatically generates an encryption key. This key is used to transfer the data securely. It guarantees that only the data subject can read the data. It is never sent to the application's server. Therefore, the service provider of the application cannot access the data. Technical Details: The browser generates a private-public key pair with javascript. Only the public key and a fingerprint (hash) are sent to the server allowing the data subject to only send the fingerprint to the controller (see later in the process). This key pair is only used for one transaction. Before submitting the form, the data subject receives an explanation of what happens with the provided data and the next steps.



Request for Access

According to Article 15 GDPR, you have the right to request a copy of your personal data from a controller, i.e. companies that process your personal data.

With this tool, you can submit a request to a controller and digitally sign it with your mobile signature.

Who has your data?

Choose Saved Company: ClearSee Ltd.

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER*: dpo@clearsee.co.uk

Who are you?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom

E-Mail*: dominik.s@dataprotector.at

Details of your Request

Optional: Customer Number or any other Identification: Customer Number: 2109381203

Optional: Any specific data you want to receive: Location data stored about me

Choose a Siganture-Service

Signature Service: Atrust Handy Signatur

Generate Encryption Key

This is your random private key, used for decrypting the received data. You must store it now securely or print it out.

Click here to see the next steps and what is being done with your data.

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Figure 6.18: Request for Access: Step 1 - Enter Data

After submitting the information, the data subject must authenticate with the chosen (eIDAS) authentication service. In this case, its Handy-Signatur (National mobile identification of Austria). After the two-factor authentication, the data subject authorises the qualified signature of the PDF.

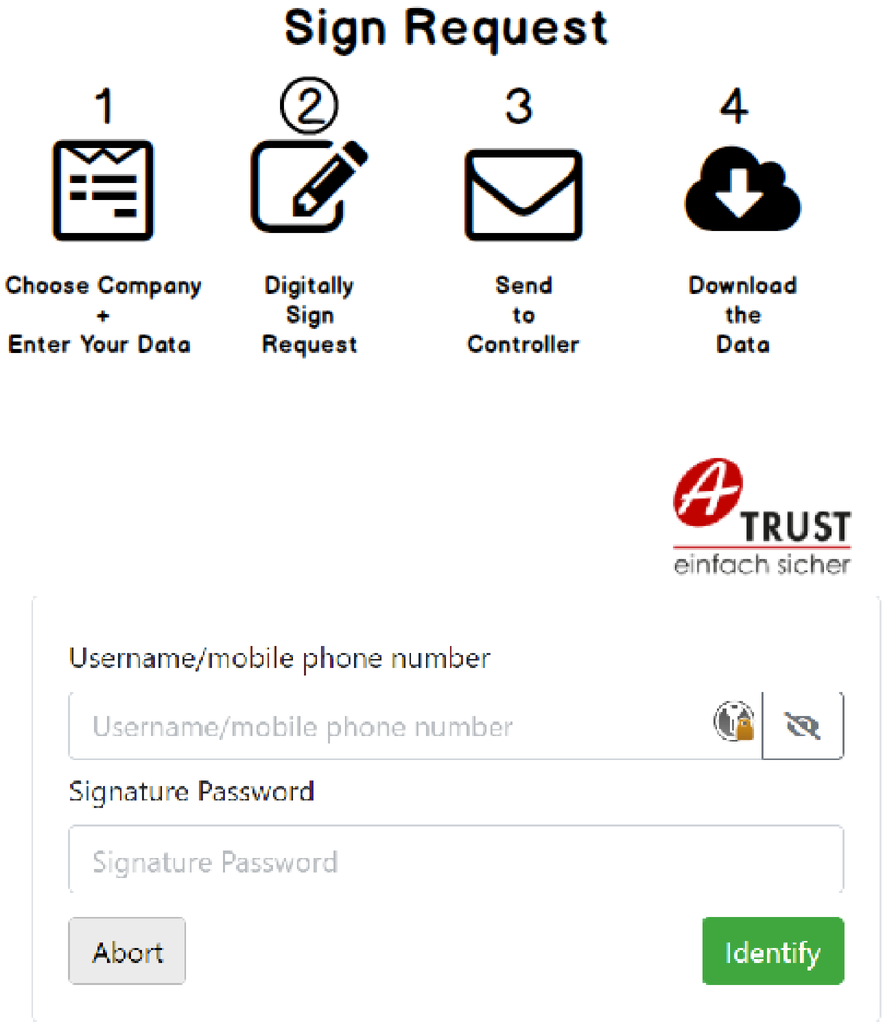


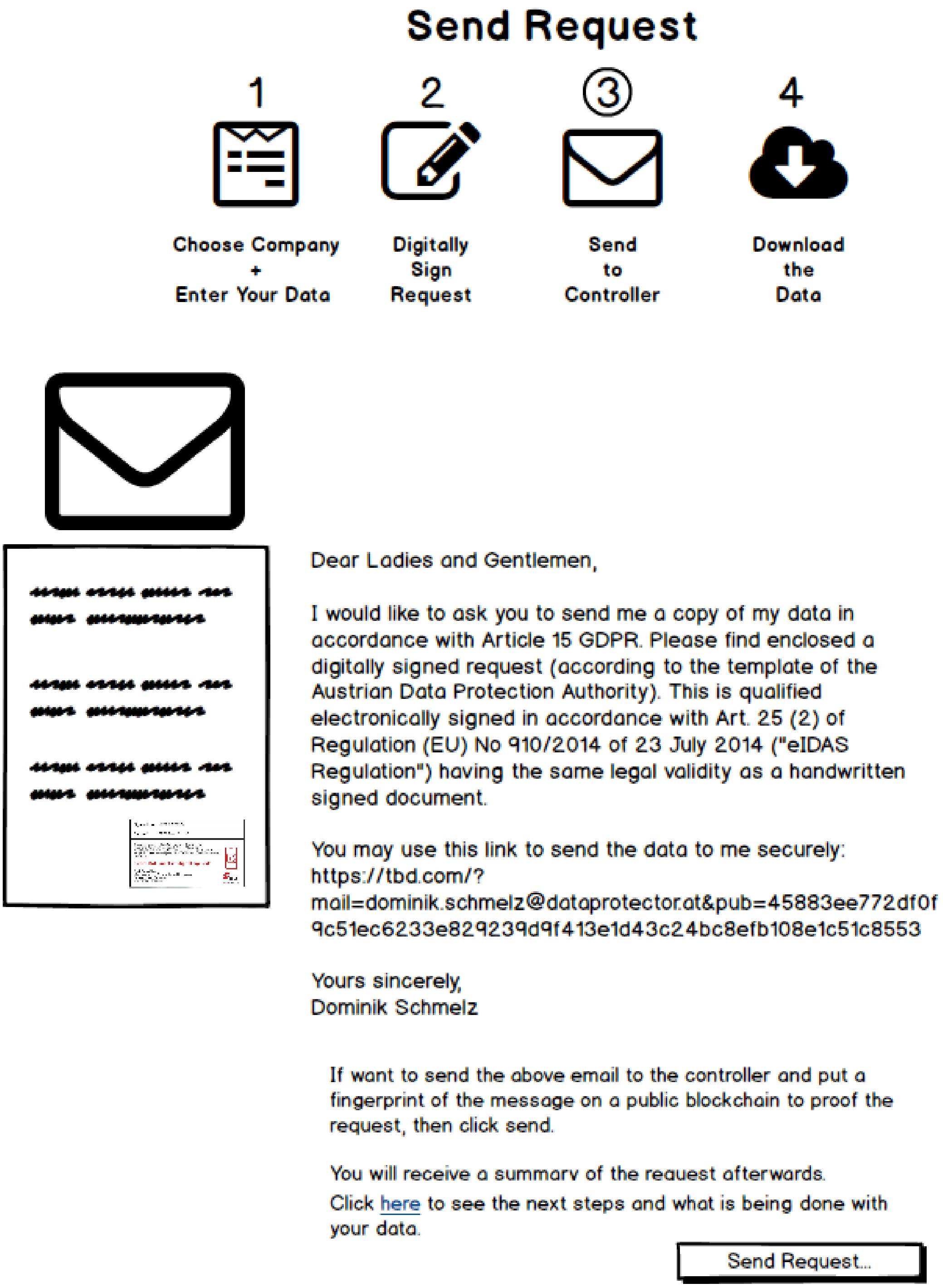
Figure 6.19: Request for Access: Step 2 - Signature

A summary is shown with the signed PDF, and the email is sent to the controller. If the data subject agrees to send the email and the PDF attached, the mail will be sent to the controller, and a fingerprint will be stored on a public blockchain.

The email is sent to the system's own SMTP Server, and the request (including the mail and PDF) is hashed and put on a public blockchain. The free OpenTimestamps service is used for this service. It timestamps the hash value, puts it in a Merkle hash tree, and stores the result on the Bitcoin blockchain. Therefore, no personal data is stored on the blockchain, and the existence of the request can be proven later.

The mail to the controller contains a link it can use to upload the data for the data subject.

Technical Detail: The link contains the email address and the fingerprint of the public key of the data subject.



Dear Ladies and Gentlemen,

I would like to ask you to send me a copy of my data in accordance with Article 15 GDPR. Please find enclosed a digitally signed request (according to the template of the Austrian Data Protection Authority). This is qualified electronically signed in accordance with Art. 25 (2) of Regulation (EU) No 910/2014 of 23 July 2014 ("eIDAS Regulation") having the same legal validity as a handwritten signed document.

You may use this link to send the data to me securely:
https://tbd.com/?mail=dominik.schmelz@dataprotector.at&pub=45883ee772df0f9c51ec6233e829239d9f413e1d43c24bc8efb108e1c51c8553

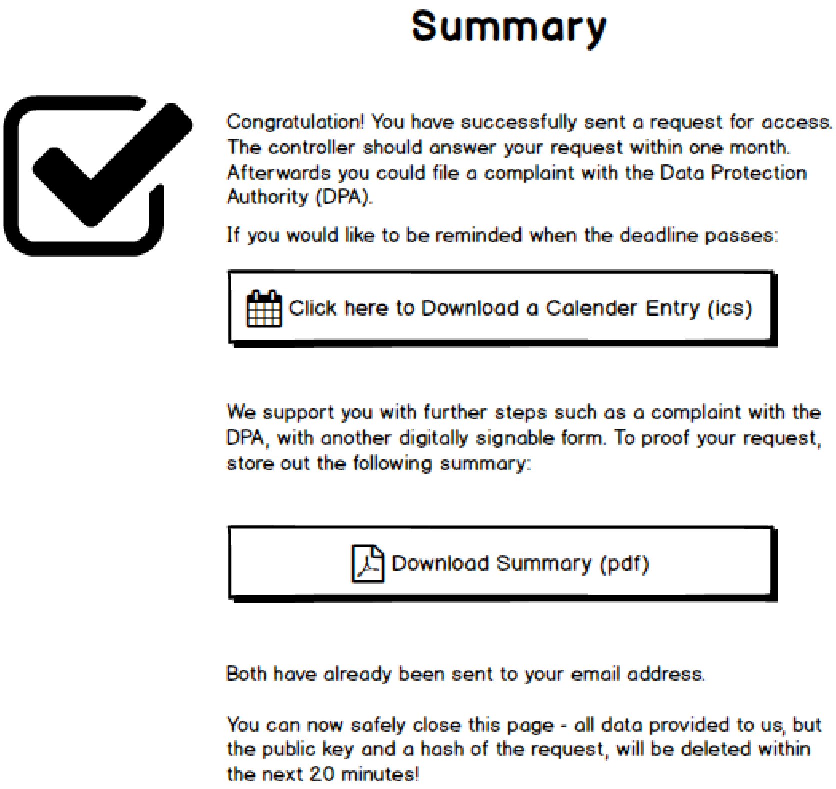
Yours sincerely,
Dominik Schmelz

If want to send the above email to the controller and put a fingerprint of the message on a public blockchain to proof the request, then click send.

You will receive a summary of the request afterwards.
Click here to see the next steps and what is being done with your data.

Figure 6.20: Request for Access: Step 3 - Send

The data subject receives a success message and confirmation via email if everything goes well. This message and the email include a summary of the request as a PDF and a calendar entry (ICS file) to be stored in the data subjects calendar software. The calendar entry reminds the data subject of the date it can file a complaint with the data protection authority if the controller has not answered the request. In case of a delay, the data subject must receive a notification from the controller; in this case, the data subject must move the calendar entry manually. It includes a link to a form to fill out the complaint information and send it again digitally signed to the Data Protection Authority.



Summary

Congratulation! You have successfully sent a request for access. The controller should answer your request within one month. Afterwards you could file a complaint with the Data Protection Authority (DPA).

If you would like to be reminded when the deadline passes:

We support you with further steps such as a complaint with the DPA, with another digitally signable form. To proof your request, store out the following summary:

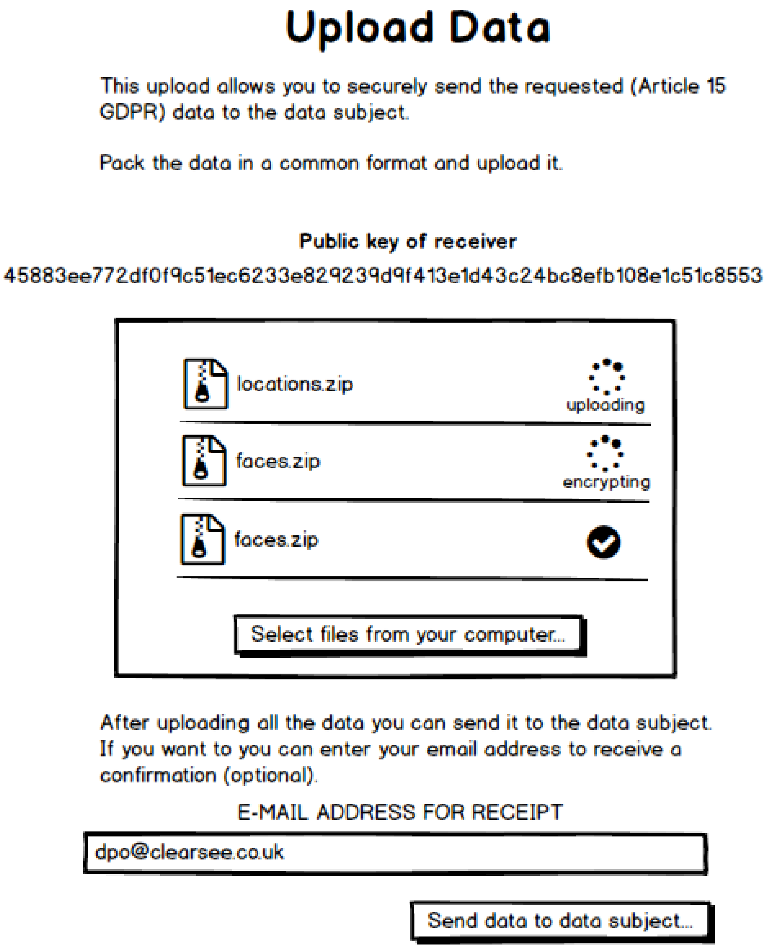
Both have already been sent to your email address.

You can now safely close this page - all data provided to us, but the public key and a hash of the request, will be deleted within the next 20 minutes!

Figure 6.21: Request for Access: Step 4 - Summary

The controller can open the link in its Browser and instantly gets the option to upload data. By either dragging and dropping files or selecting files with the file, browser files can be added. These are automatically encrypted. Technical Details: The web link contains the fingerprint of the public key. This is used to retrieve the public key that is then used by the Browser to encrypt the files with hybrid encryption, meaning that the data is symmetrically encrypted, and the symmetric key is encrypted with the asymmetric public key achieving faster encryption.

Optionally the controller can enter an email address to receive a summary of the uploaded files.



Upload Data

This upload allows you to securely send the requested (Article 15 GDPR) data to the data subject.

Pack the data in a common format and upload it.

Public key of receiver

45883ee772df0f9c51ec6233e829239d9f413e1d43c24bc8efb108e1c51c8553

After uploading all the data you can send it to the data subject. If you want to you can enter your email address to receive a confirmation (optional).

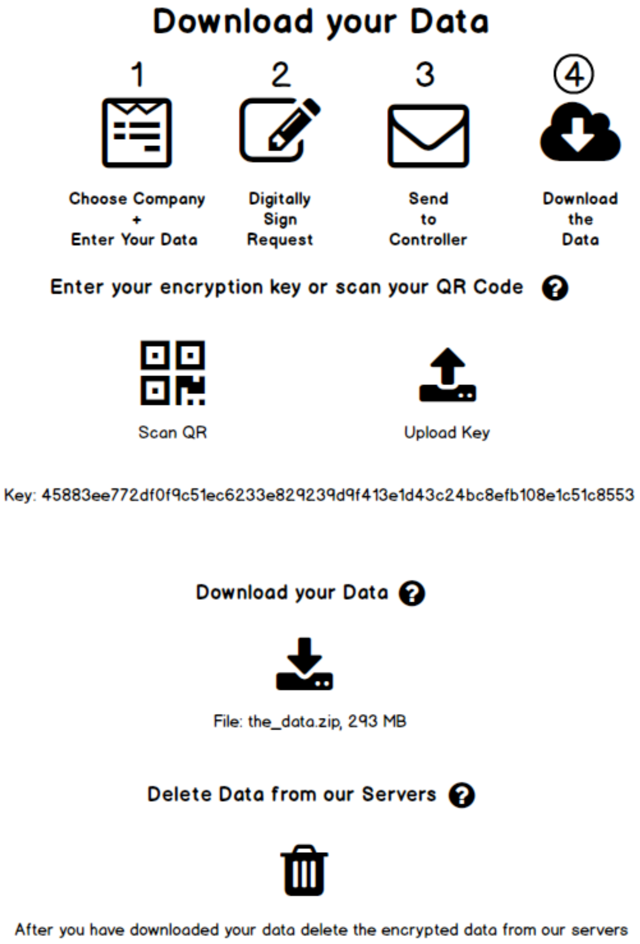
E-MAIL ADDRESS FOR RECEIPT

dpo@clearsee.co.uk

Figure 6.22: Request for Access: Step 4.1 - Upload

After the data has been uploaded by the controller, the data subject receives a notification. This notification links to a webpage that allows the data subject to upload or scan its private key. Then the data is decrypted and can be downloaded. When the download has finished and the decryption is successful, a deletion request can be submitted to the server of the application. This concludes the data transfer, and no further data is stored on the application servers.

Technical Detail: The encrypted data is retrieved from the application server's binary storage; The private key is validated against the fingerprint, and the decryption is done in the browser.



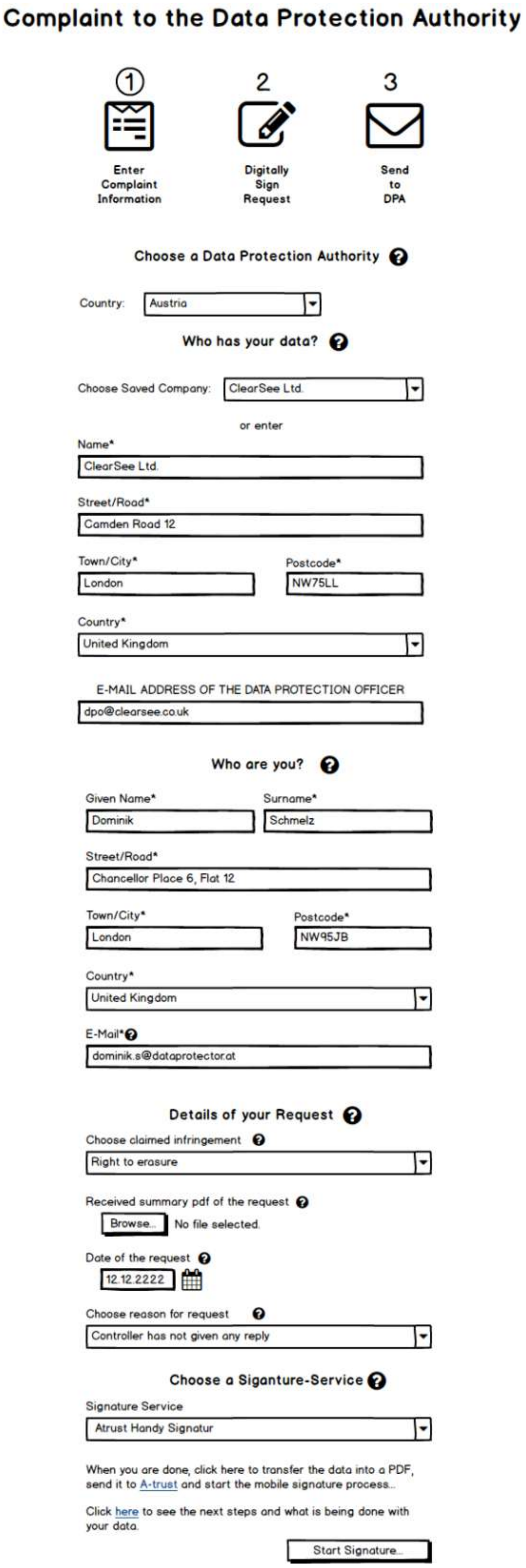
Key: 45883ee772df0f9c51ec6233e829239d9f413e1d43c24bc8efb108e1c51c8553

File: the_data.zip, 293 MB

After you have downloaded your data delete the encrypted data from our servers

Figure 6.23: Request for Access: Step 5 - Download

To file a complaint, the data subject can click the link provided in the calendar entry. The data subject can choose a data protection authority it wants to address. It must specify the details of the request. These have been derived from the Austrian data protection authorities' complaint form. The summary received via email can be attached to the complaint to prove the request to the authority. Many fields of the complaint form are automatically filled out (such as type of signature, type of contact etc.).



Complaint to the Data Protection Authority

Choose a Data Protection Authority ?

Country: Austria

Who has your data? ?

Choose Saved Company: ClearSee Ltd.

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER: dpo@clearsee.co.uk

Who are you? ?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom

E-Mail* ?: dominik.s@dataprotector.at

Details of your Request ?

Choose claimed infringement ?: Right to erasure

Received summary pdf of the request ?: No file selected.

Date of the request ?: 12.12.2222

Choose reason for request ?: Controller has not given any reply

Choose a Siganture-Service ?

Signature Service: Atrust Handy Signatur

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Click here to see the next steps and what is being done with your data.

Figure 6.24: Request for Access: Optional Step 6 - Complaint

6.5 Qualitative Evaluation of the Case Studies

The previous sections illustrated two GDPR processes using BPMN, used Delphi expert panels to find criteria for applications that would improve activities in these processes, and ranked them in order of importance.

This section will make use of the processes and criteria questions generated in the previous sections to put two prototypical applications to the test.

6.5.1 Design

The hypothesis to be tested in this evaluation was "The case studies shown support activities in the GDPR erasure and access process". An expert survey was conducted to answer the question of whether they support which activities and, if so, how strongly they support them. The quantitative expert survey was conducted through an online survey using selected experts. Compared to qualitative methods, the quantitative method allows for accurate testing of hypotheses [Sti05, p. 91]. For this quantitative study, an interview in the form of a questionnaire was used. Again, an anonymous question-based interview was chosen [May09, p. 100].

For scientific compliance, each participant had to agree to a scientific evaluation of the information provided. The participants were also always able to cancel the survey and were not forced to give answers in any way.

The survey, on the one hand, explained the designed processes and, on the other hand, explained the case studies (see Appendix E for the full survey form). This was done in order to have the exact same explanation provided to every expert.

Each expert was provided with criteria questions regarding each activity. Then the experts were asked to rate the support of the application for the according criteria.

Population For this purpose, seven domain experts (n=7) in the field were selected as in the previous qualitative research. Since both data protection and product development aspects are to be covered, experts in both fields have been interviewed. The experts needed to be able to understand BPMN, the application prototypes and to assess criteria questions.

Data Collection The data was collected by asking questions and rating each activity. This was done with the help of pictures depicting user interfaces and an explanation of the steps within the user interface. Also, the processes which activities were asked about were depicted in diagrams. For the survey creation and answering, the online platform SurveyMonkey [4] was used because of its ease of use and GDPR compliance [Sur21].

[4] https://www.surveymonkey.com

Questions The questions were divided into two sections: erasure and access. Each section explained the according process and application. Each activity visible in Figures 6.2 and B.5 were discussed.

Each activity was explained, and the according criteria questions (see Table 6.5 for an example) were asked. The experts were then asked to rate the supporting features within the application for the activity.

The options to rate an evaluation criterion were limited by a linear rating with five options according to the Likert scale [JKCP15]. The following options were given (arranged in a logical order):

1. Not supportive
2. Little supportive
3. Supportive
4. Well supportive
5. Very supportive

The options were given points (1-5) but were not shown to the experts.

6.5.2 Analysis

All criteria question responses were added up for each question by using a weighted average, resulting in an 'Application Support Score' for each question. The ranking score for each question, also called the 'Question Importance Score', from the last Delphi round was taken as a multiplier or weight for the importance of the questions and, therefore, the result. The 'Question Importance Score' values where then used for the weighted average calculation of the 'Application Support Score' resulting in a final 'Weighted Score' for each activity. Table 6.6 shows the results of the process activities in the process 'request for erasure'.

Table 6.6: Scores of support of process activities in the process request for erasure

Process Activity	Weighted Score
RFE.S_1_1	4,1
RFE.S_1_2	3,6
RFE.S_1_3	4,0
RFE.S_2	4,5
RFE.S_6	3,7
RFE.S_7	4,0
RFE.S_8	3,8
RFE.S_9	4,3
RFE.S_4	4,4
RFE.S_5	4,1

Table 6.7 shows the results of the process activities in the process 'request for access'.

Table 6.7: Scores of support of process activities in the process request for access

Process Activity	Weighted Score
RFA.S_1_1	4,3
RFA.S_1_2	3,9
RFA.S_1_3	3,9
RFA.S_2	4,2
RFA.S_6	3,3
RFA.S_7	3,9
RFA.S_8	3,5
RFA.S_9	3,5
RFA.S_4	3,8
RFA.S_5	3,6

Interpretation The results show that some process activities are more supported than others. The average score for a request for access is lower (3,8) than the average score for the erasure activities (4,0).

Figure 6.25 shows the process of the request for erasure with a focus on the data subject coloured according to the supporting features. The green highlighted activities are rated above average. The light-red coloured activities are below average.

Figure 6.26 shows the same illustration for the 'request for access' process.

These figures show that certain process activities were supported in both applications. Namely creation processes (*S_1_1*), proofing the identity (*S_2*) and filing a complaint.

Handling of results such as positive, delayed answers or rejections is not well supported by both applications. Only the activity 'Check Partial Erasure Explanation' (*S_9*) was rated highly. This is mostly due to the option of filing a complaint and the explanation of the terminology.

Filing a complaint was highly rated for both processes. The complaint result was low rated because of the missing possibility of further remedies.

The highest rated questions were regarding fault tolerance, predictability, trustworthiness, security and usability.

6.5.3 Conclusions

The presented case studies were reviewed by seven experts in terms of their support for GDPR processes from the perspective of data subjects. The applications were prototypes of two web applications that support blockchain technology. Although features that do

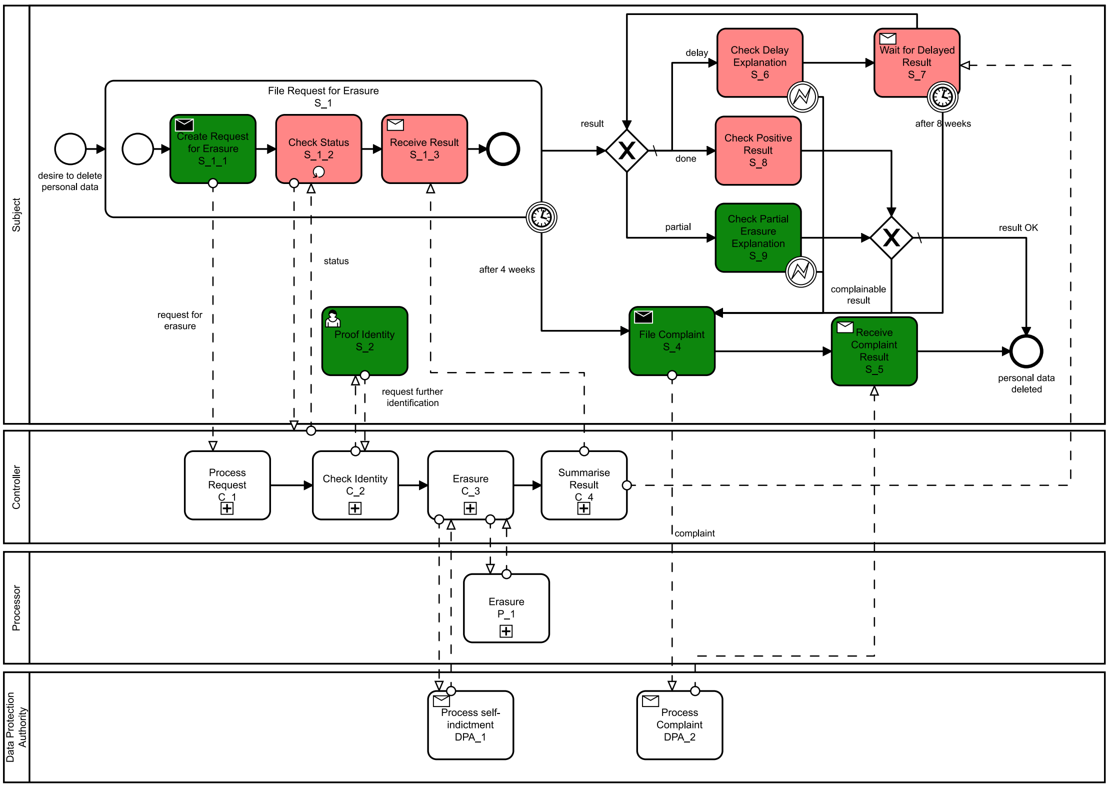


Figure 6.25: Process of the request for erasure with a focus on the data subject coloured according to the supporting features

not directly relate to the blockchain technology were also evaluated, features that result directly from the use of the blockchain were evaluated positively.

Specifically, blockchain technology helped to implement a verifiable and demonstrable transmission of requests and secure data transfer between the data controller and the data subject.

This step concludes the assessment process for legal process support technologies defined in Chapter 5. It has been shown how laws can be represented as processes; the existence and separation of problems within the processes can be assessed; assessment criteria can be created regarding the supporting properties of applications for activities; and finally, applications can be assessed against these assessment criteria.

6.6 Discussion

Chapter 4 showed that blockchain technology could be difficult to use in a data-protected way because of its fundamental properties, such as openness and immutability. The case studies showed how blockchain technology could be used in a meaningful way to support

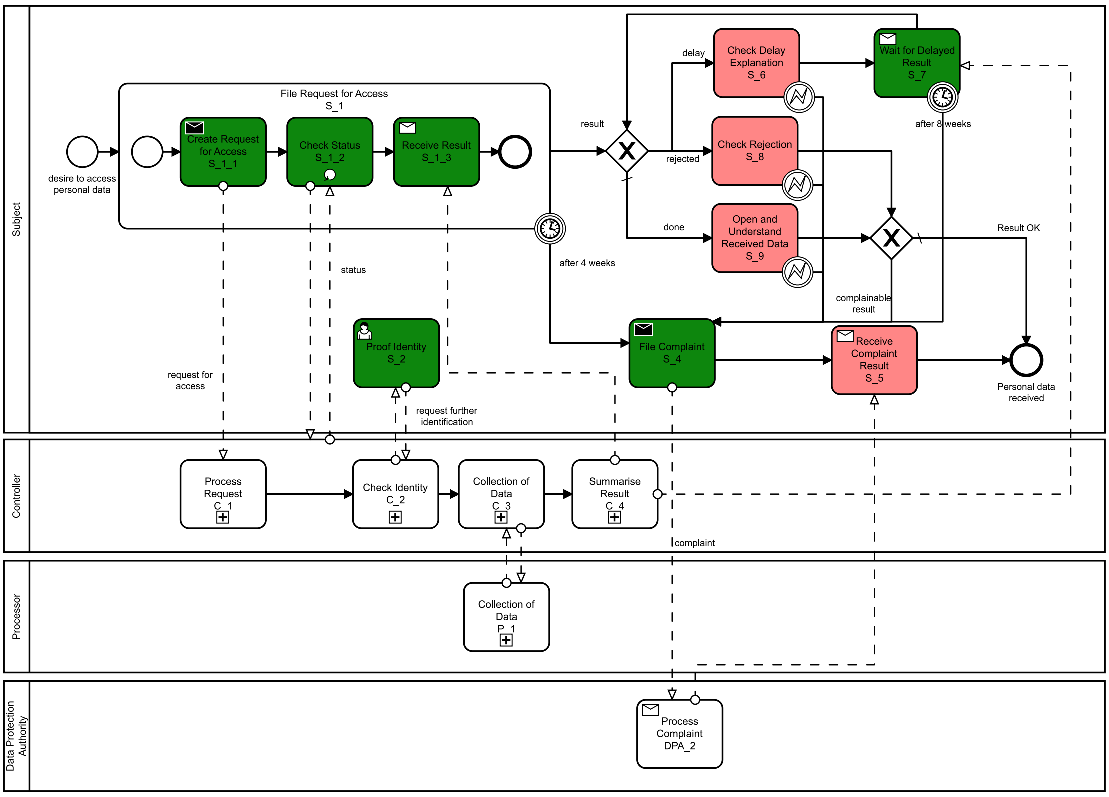


Figure 6.26: Process of the request for access with a focus on the data subject coloured according to the supporting features

certain data protection properties answering the question: “Can blockchain technology be used in applications to support data protection according to the GDPR?”.

They used blockchain technology in a way that specifically exploited the properties of the blockchain. These features allowed the data subject to present strong evidence that a request for deletion or access was made. They also allowed a secure, end-to-end encrypted data transfer, which is essential for the privacy of the data subject with regard to the data to be transferred. The applications have shown how blockchain technologies can be used to design applications that allow a data subject to enforce his rights without further compromising his privacy, answering the question: “Can blockchain-based applications help empower data subjects?”.

Critically, a trusted third party, such as the data protection authority, could take on the role of the blockchain. For example, this independent party could provide a notary service or even an application service so that proof of application can be provided. As long as this is not the case, an independent third party has to take over this role. In the case of the applications shown, this role is taken by the blockchain. Through the mechanisms shown in the previous chapters, the blockchain itself can take on this role.

Thanks to a large number of participants and sophisticated algorithms, it is able to store data in a fairly unchangeable, highly distributed manner and is, therefore, resilient to attacks. These properties allow citizens to remove not only banks from a central role of trust but also other entities - all in order to be able to claim rights that have already been granted.

CHAPTER 7

Conclusion and Future Work

Privacy can be enhanced by technology, but at the same time, the use of certain technologies can compromise privacy and data protection.

Blockchain technology was created to make centralised trust obsolete and to promote decentralised structures in which every individual can participate in a system to transfer information and value without restrictions and minimal trust. Blockchain technology is built on the principles that, if used without reflection, endanger the privacy of participants but also the privacy of those referenced within transferred data.

This work bridges the technical-legal field of privacy with and despite blockchain technology. It connects the field of technical data protection as defined by the GDPR and the legal data protection aspects of blockchain technology. Based on this, privacy issues of Blockchain technology and possible solutions were analysed in a risk-based manner. For this purpose, a process was designed which uses a novel risk-based approach based on established standards.

Furthermore, the thesis shows blockchain technology as a solution to the privacy issues of the affected parties. It outlines a process that can be used to design and evaluate legal processes supporting applications. This process is used to evaluate two different applications that support data subjects' rights. For this purpose, the problem areas are elicited and quantified from the point of view of the data subjects on the basis of a business process map of the GDPR utilising a quantitative survey. These are tried to solve or improve employing technology using blockchain. In order to validate these prototypically designed solutions, a criteria catalogue was created based on the same business process of the GDPR, and the applications are evaluated by means of an expert survey.

The research has shown that, on the one hand, many technical solutions previously thought of as solutions for data protection issues in the field of blockchain technology do not meet expectations. On the other hand, the research has shown that blockchain-based applications can support data subjects to achieve their rights without further compromising their privacy.

APPENDIX A

Appendix: Risk-based Assessment Template

The following table shows a template allowing risk registration based on the rights and obligations defined in the GDPR.

Obligations	**GDPR Art.**	**Risk and Impact Description**	**Impact Evaluation**	**Probability Evaluation**	**Risk Priority**	**Mitigation**	**Residual Risk**
Records and transparency	30						
Technical and organisational obligations	32						
Data protection impact assessment	35						
Data-breach notification	34						
Appointment of a data protection officer	37						
Rights							
Right of access by the data subject	15						
Right to rectification	16						
Right to erasure	17						
Right to restriction of processing	18						
Notification obligation	19						
Right to data portability	20						
Right to object	21						
Automated decision-making & profiling	22						

Table A.1: Template to assess risks based on rights and obligations defined by the GDPR

APPENDIX B

Appendix: Models of GDPR Processes

The following figures document the created business process diagrams during the course of research. The diagrams were created in Business Process Model and Notation Version 2.0 with the help of Camunda's BPMN software 'Camunda Modeler'[1]. The source files will be put in the TU Wien research data archives. A summary of the processes and subprocesses can be found in Table B.1 and B.2.

B.1 Overview of Modelled Erasure Processes

[1] https://bpmn.io

Table B.1: Summary of the processes and subprocesses involved in the request for erasure

Process	Subprocess	Description	Actors	Referred GDPR Articles
Erasure		Erasure process with focus on data subject	Subject, Controller, Processor, Data Protection Authority	Art. 17
Erasure		Erasure process with focus on controller	Subject, Controller, Processor, Data Protection Authority	Art. 17, Art. 12
Erasure	Check Grounds	Subprocess of controller erasure to check legal grounds	Controller	Art. 17, Art. 8
Erasure	Check Necessity of Processing	Subprocess of controller erasure to check necessity of processing still exists	Controller	Art. 17
Erasure	Check Identity	Subprocess of controller erasure to check the identity of the subject	Subject, Controller	Art. 17
Erasure	Erasure	Subprocess of controller erasure to delete the correct data if possible	Controller, Processor, Data Protection Authority	Art. 17, Art. 12

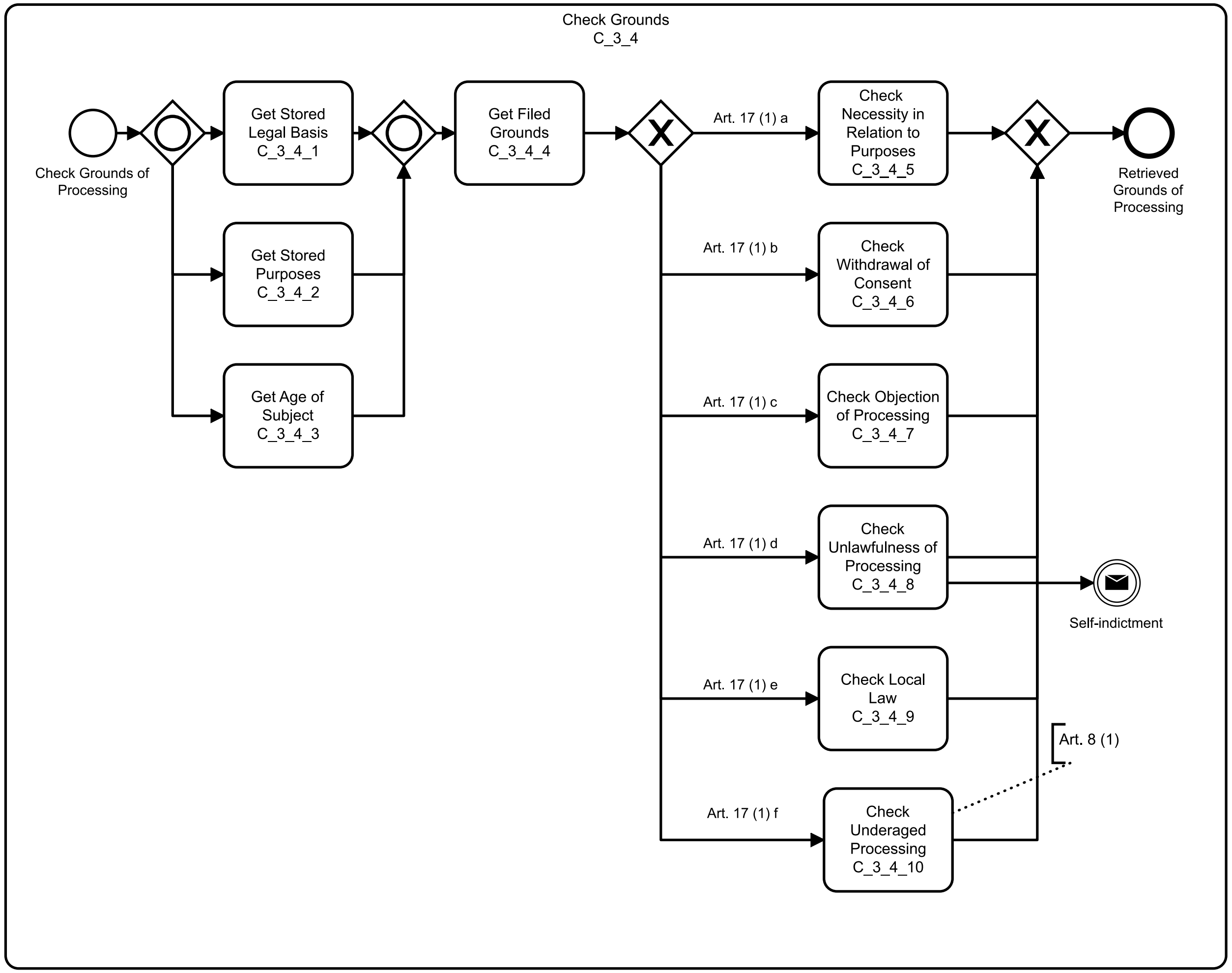


Figure B.1: Subprocess of the erasure process that checks the legal grounds

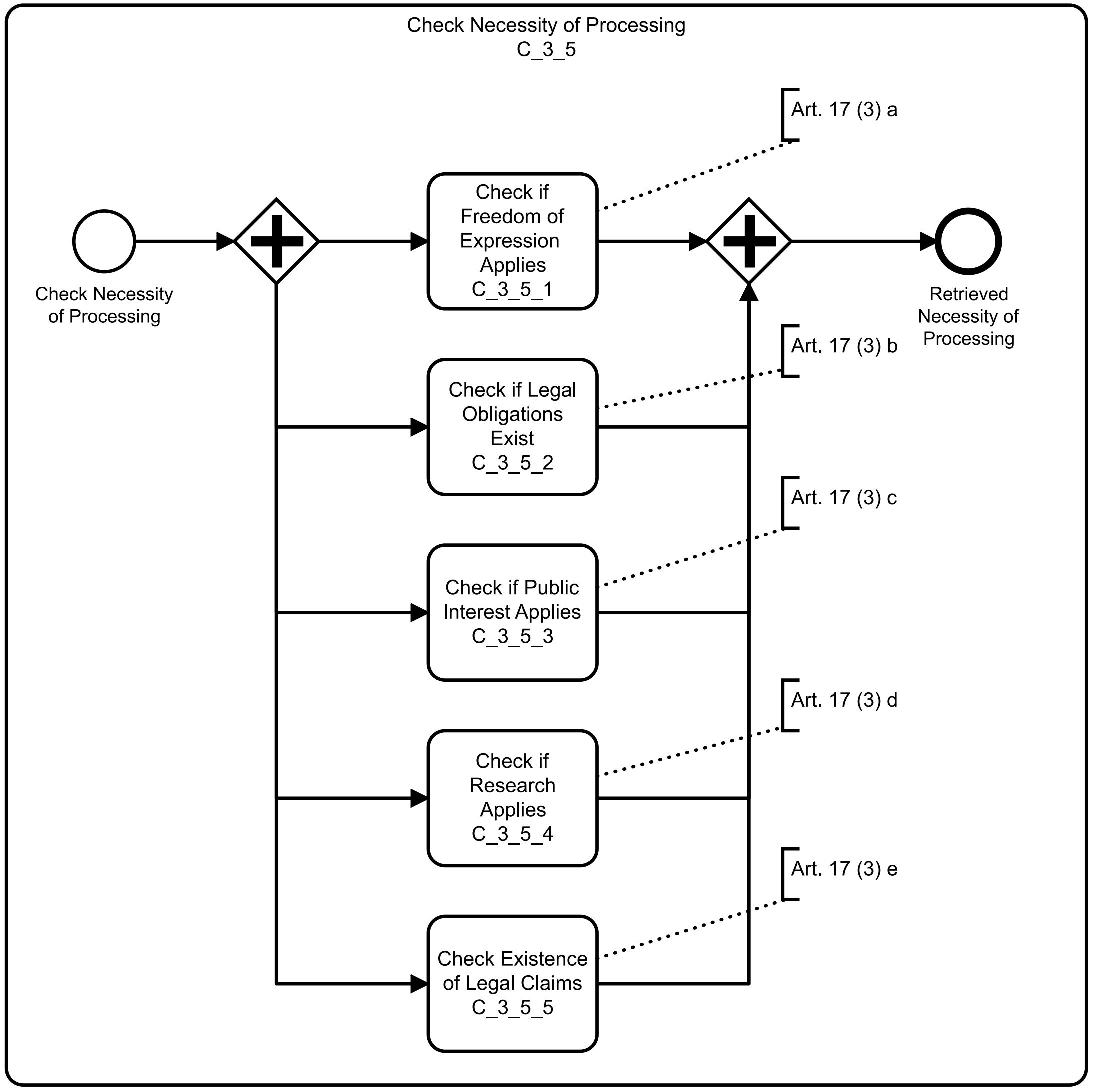

Check Necessity of Processing
C_3_5
Check Necessity of Processing
Check if Freedom of Expression Applies C_3_5_1
Check if Legal Obligations Exist C_3_5_2
Check if Public Interest Applies C_3_5_3
Check if Research Applies C_3_5_4
Check Existence of Legal Claims C_3_5_5
Art. 17 (3) a
Art. 17 (3) b
Art. 17 (3) c
Art. 17 (3) d
Art. 17 (3) e
Retrieved Necessity of Processing

Figure B.2: Subprocess of the erasure process that checks the necessity of a deletion

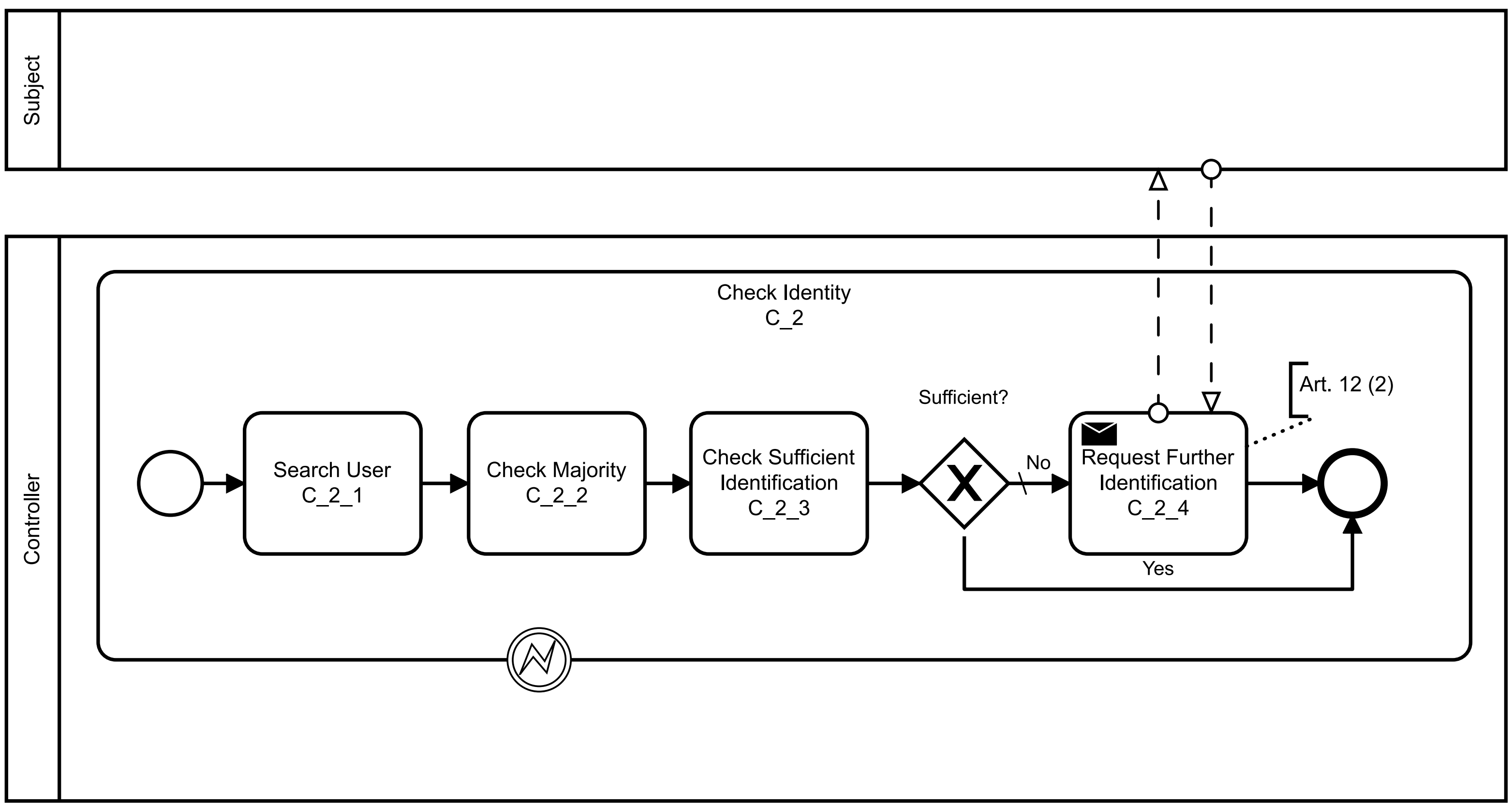


Figure B.3: Subprocess of the erasure process that checks the identity of the data subject

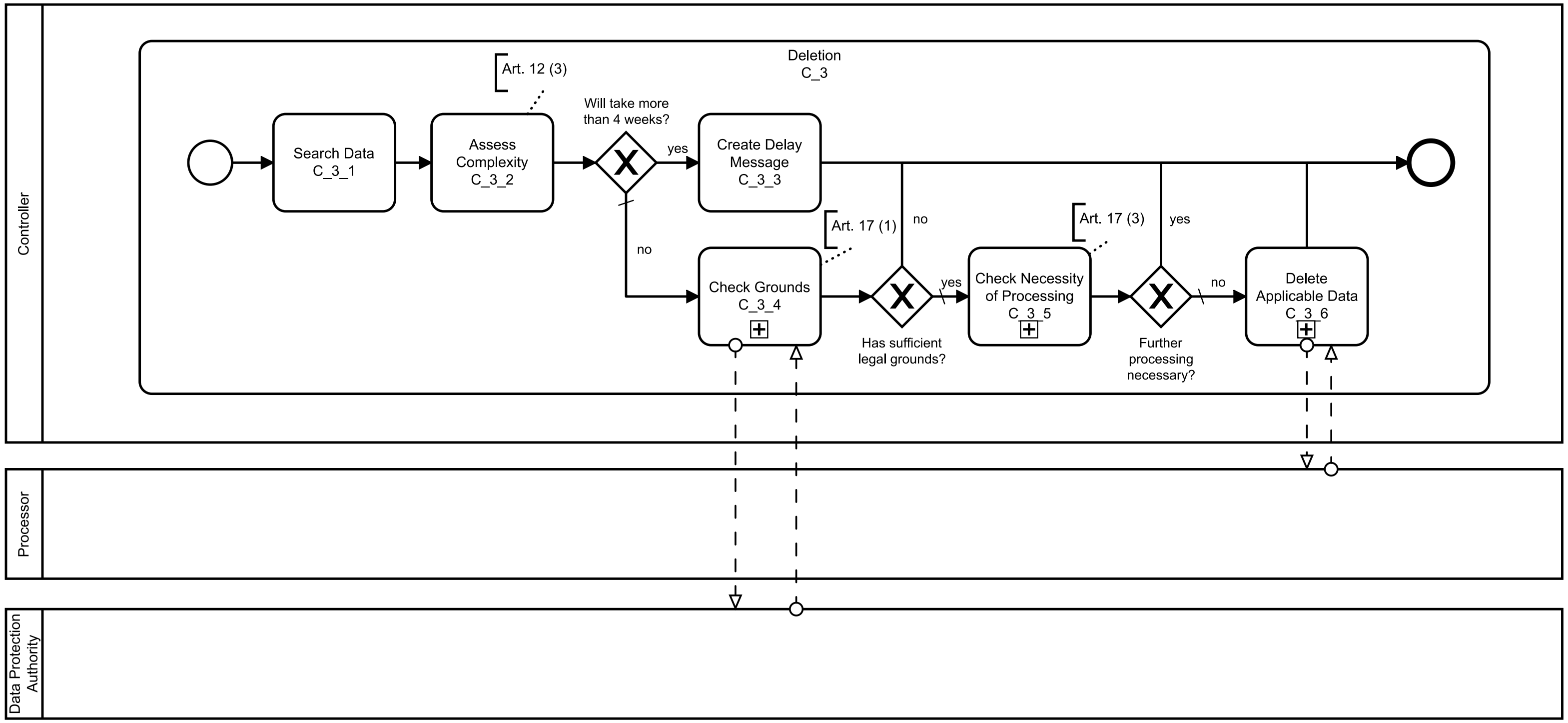


Figure B.4: Subprocess of the erasure process that does the actual deletion

Table B.2: Summary of the processes and subprocesses involved in the request for access

Process	Subprocess	Description	Actors	Referred GDPR Articles
Access		Access process with focus on data subject	Subject, Controller, Processor, Data Protection Authority	Art. 15
Access		Access process with focus on controller	Subject, Controller, Processor	Art. 15, Art. 12
Access	Check Identity	Subprocess of the controller access process to check the identity of the subject (Similar to erasure process)	Subject, Controller	Art. 15
Access	Collection	Subprocess of controller collecting the information needed to answer the request	Controller, Processor	Art. 15, Art. 12

B.2 Overview of Modelled Access Processes

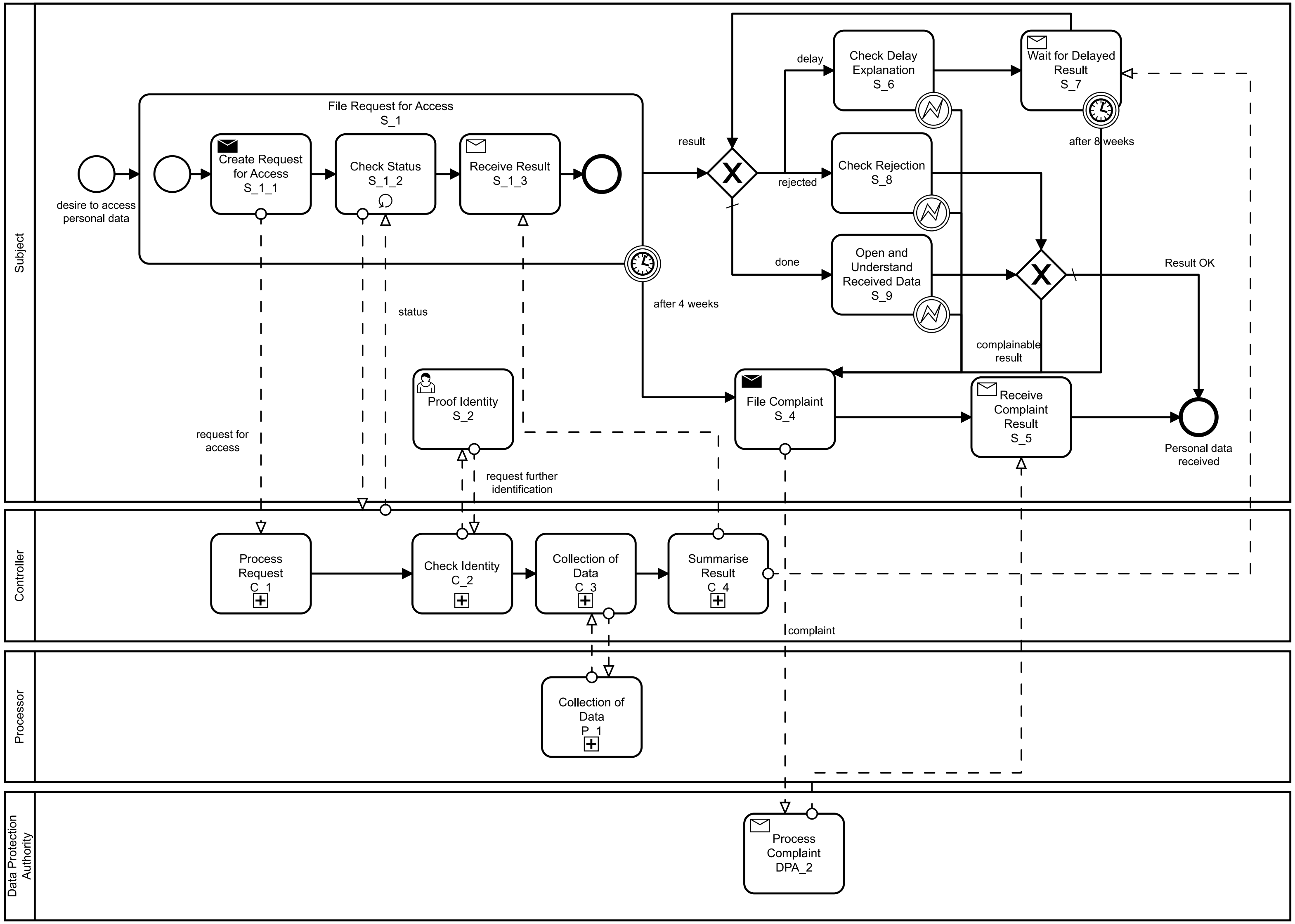


Figure B.5: Process of the request for access with a focus on the data subject

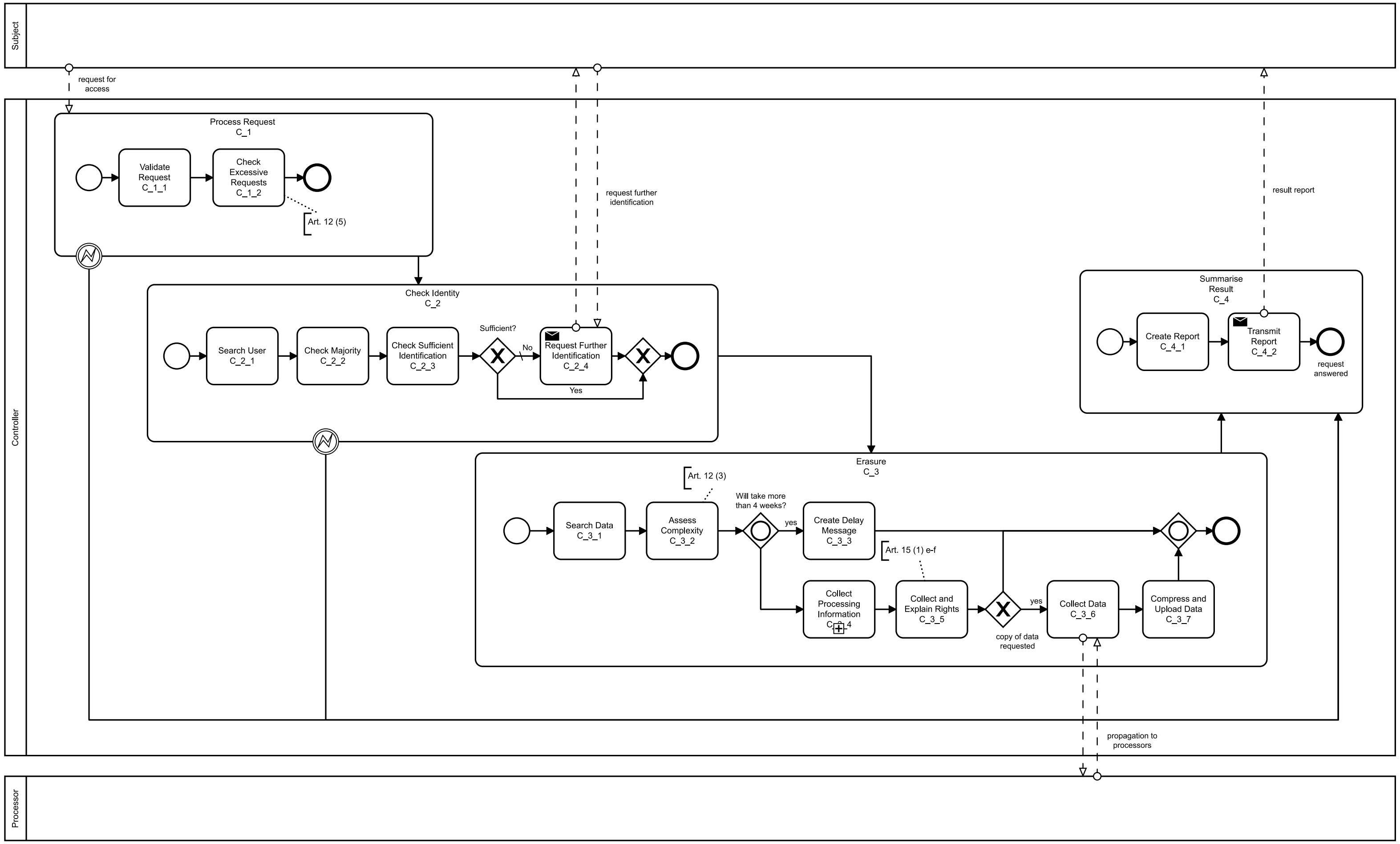


Figure B.6: Process of the request for access with a focus on the controller

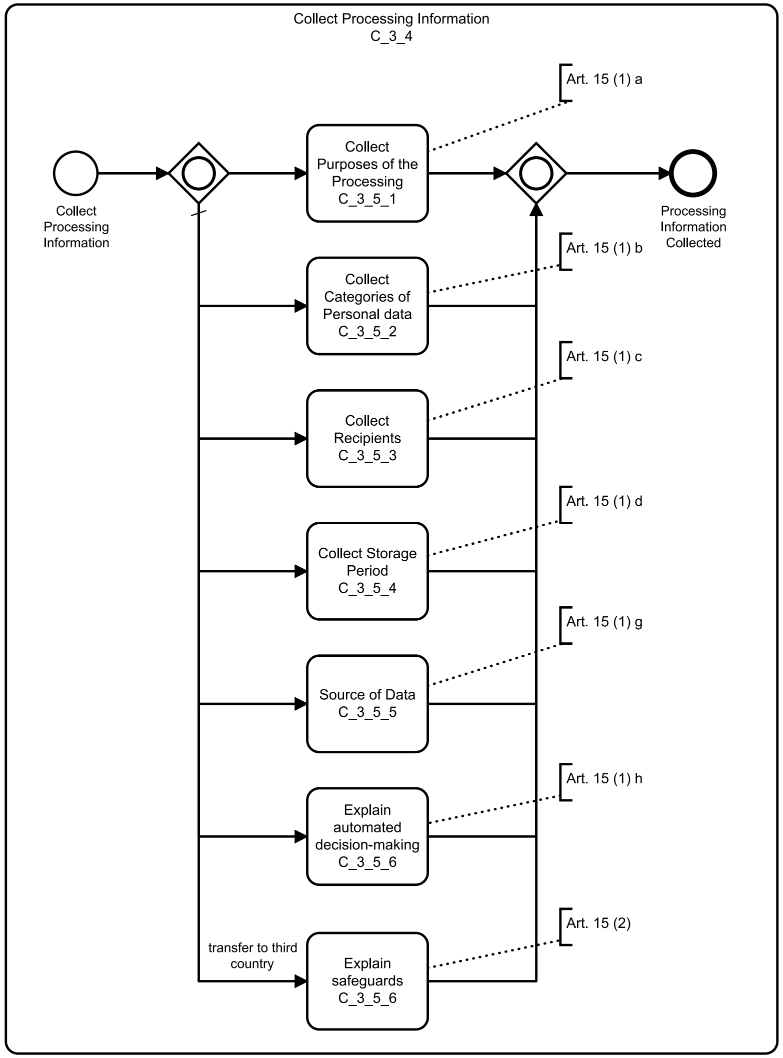

Collect Processing Information
C_3_4
Art. 15 (1) a
Collect Processing Information
Collect Purposes of the Processing C_3_5_1
Processing Information Collected
Art. 15 (1) b
Collect Categories of Personal data C_3_5_2
Art. 15 (1) c
Collect Recipients C_3_5_3
Art. 15 (1) d
Collect Storage Period C_3_5_4
Art. 15 (1) g
Source of Data C_3_5_5
Art. 15 (1) h
Explain automated decision-making C_3_5_6
Art. 15 (2)
transfer to third country
Explain safeguards C_3_5_6

Figure B.7: Subprocess of the access process collecting the information needed to answer the request

APPENDIX C

Appendix: Survey of Issues in Data Protection Processes

C.1 Questions

Table C.1 lists the questions asked in the survey.

Table C.1: Questions of the survey

#	**Original Question**	**English Question**	**Original Answer Options**	**English Answer Options**
1	Ich stimme den genannten Bedingungen dieser Umfrage zu.	I agree to the stated terms of this survey.	Zustimmen Ablehnen	Agree Decline
2	In welchem Land leben Sie momentan?	In which country do you currently live?	Deutschland Österreich Schweiz Keine der oben genannten	Germany Austria Switzerland None of the above
3	Wie alt sind Sie?	How old are you?	Unter 18 18-20 21-29 30-39 40-49 50-59 Über 60	Under 18 18-20 21-29 30-39 40-49 50-59 Over 60

4	Was ist der höchste Bildungsgrad, den Sie bisher erlangt haben?	What is the highest level of education you have obtained so far?	Weniger als Grundschule Grundschule Hauptschulabschluss Realschulabschluss bzw. Mittlere Reife Fachhochschulreife bzw. Abitur Bachelor Master bzw. Magister Diplom Promotion	Less than elementary school Elementary school Secondary school diploma Realschulabschluss or Mittlere Reife (intermediate school leaving certificate) Advanced technical college certificate or high school diploma Bachelor's degree Master or Magister Diploma Doctorate
5	Sind Sie männlich oder weiblich?	Are you male or female?	Männlich Weiblich	Male Female
6	Haben Sie einmal einen Antrag auf Löschung oder Auskunft nach DSGVO gestellt?	Have you ever submitted an erasure or access request under the GDPR?	Ja Nein Nein, ich wusste nicht, dass ich das Recht dazu habe	Yes No No, I did not know that I had the right to do so
7	Wie oft haben Sie bereits einen Löschantrag (Art. 17 DSGVO) gestellt?	How many times have you already submitted an erasure request (Art. 17 GDPR)?	Noch nie Einmal Mehrmals	Never Once Several times

8	Bei meinem letzten Löschantrag fand ich die folgenden Schritte...	In my last erasure request, I found the following steps...	sehr einfach einfach ging so schwer sehr schwer war nicht der Fall Auffinden und Ausfüllen des Antragsformulars Identifikation gegenüber der Firma (zB Ausweis scannen) Beschwerde bei der Datenschutzbehörde Verstehen der Antwort der Firma Nachricht über Verzögerung verstehen Verstehen warum nicht alle meine Daten gelöscht wurden Antwort der Beschwerde bei der Datenschutzbehörde verstehen Erhalten des Status des Antrags	very simply simply ok difficult very difficult was not the case Finding and filling in the application form Identification to the company (e.g. scan ID card) Complaining to the data protection authority Understanding the company's response Understanding message about delay Understand why not all my data was deleted Understand the response of the complaint to the data protection authority Getting the status of the request

9	Wie haben Sie den letzten Löschantrag gestellt?	How did you submit the last erasure request?	In einem Online-Formular der Firma Formlos per E-Mail oder Post Mittels eines Anrufes Mittels eines Formulars per E-Mail oder Post Sonstiges (bitte angeben)	In an online form of the company Informally by email or mail By means of a phone call By means of a form sent by e-mail or post Other (please specify)
10	Wie haben Sie die Antwort des letzten Löschantrages bekommen?	How did you receive the response of the last erasure request?	Online auf der Webseite der Firma Per E-Mail Mittels eines Anrufes Per Post Sonstiges (bitte angeben)	Online on the company's website By e-mail By phone call By mail Other (please specify)
11	Wie haben Sie sich gegenüber der Firma ausgewiesen?	How did you identify yourself to the company?	Scan oder Fotos eines Ausweises Login auf der Webseite (keine weitere Identifikation gefordert) Digitaler Ausweis oder Signatur (Handy Signatur, Bürgerkarte etc.) Sonstiges (bitte angeben) Ich musste mich nicht ausweisen	Scan or photos of an identification document Login on the website (no further identification required) Digital ID or signature (mobile signature, citizen card, etc.) Other (please specify) No identification required

12	Wie lange haben Sie auf die Antwort gewartet?	How long did you wait for the response?	Weniger als 7 Tage Weniger als 4 Wochen Mehr als 4 Wochen Ich habe keine Antwort erhalten	Less than 7 days Less than 4 weeks More than 4 weeks I have not received a response
13	Wie oft haben Sie einen Antrag auf Auskunft (Art. 15 DSGVO) gestellt?	How many times have you made a request for information (Art. 15 GDPR)?	Noch nie Einmal Mehrmals	Never Once Several times

14	Bei meinem letzten Auskunftsantrag fand ich die folgenden Schritte...	In my last request for access, I found the following steps...	sehr einfach einfach ging so schwer sehr schwer war nicht der Fall Auffinden und Ausfüllen des Antragsformulars Identifikation gegenüber der Firma (zB Ausweis scannen) Empfang der Daten Öffnen und Verstehen der Daten Beschwerde bei der Datenschutzbehörde Nachricht über Verzögerung verstehen Antwort der Beschwerde bei der Datenschutzbehörde verstehen Erhalten des Status des Antrags	Very simple simply ok difficult very difficult was not the case Finding and filling in the application form Identification to the company (e.g. scan ID card) Receiving the data Opening and understanding the data Complaining to the data protection authority Understanding message about delay Understanding the response to the complaint to the data protection authority Receiving the status of the request

15	Wie haben Sie den letzten Antrag auf Auskunft gestellt?	How did you make the last request for access?	In einem Online-Formular der Firma Formlos per E-Mail oder Post Mittels eines Anrufes Mittels eines Formulars per E-Mail oder Post Sonstiges (bitte angeben)	In an online form of the company Informally by e-mail or mail By means of a phone call By means of a form sent by e-mail or post Other (please specify)
16	Wie haben Sie die Daten erhalten?	How did you obtain the data?	Als PDF Als strukturierte Datendatei (zB Excel, CSV oder ähnliches) Als Brief bzw. ausgedruckt Telefonisch Sonstiges (bitte angeben) Ich habe die Daten nie erhalten	As a PDF As a structured data file (e.g. Excel, CSV or similar) As a letter or printed out By telephone Other (please specify) I never received the data

17	Wie haben Sie sich gegenüber der Firma ausgewiesen?	How did you identify yourself to the company?	Scan oder Fotos eines Ausweises Login auf der Webseite (keine weitere Identifikation gefordert) Digitaler Ausweis oder Signatur (Handy Signatur, Bürgerkarte etc.) Sonstiges (bitte angeben) Ich musste mich nicht ausweisen	Scan or photos of an ID card Login on the website (no further identification required) Digital ID or signature (mobile signature, citizen card, etc.) Other (please specify) I was not required to provide identification
18	Wie lange haben Sie auf die Antwort gewartet?	How long did you wait for the response?	Weniger als 7 Tage Weniger als 4 Wochen Mehr als 4 Wochen Ich habe keine Antwort erhalten	Less than 7 days Less than 4 weeks More than 4 weeks I did not receive a reply

19	Letzte Frage: Reihen Sie bitte, was Ihnen besonders wichtig ist bei Anträgen	Last question: Please tell us what is particularly important to you in the case of requests for information.	dass der Antrag verständlich ist dass meine Daten sicher sind dass wenig Daten abgefragt werden dass er formlos sein kann dass ich ihn von meinem Smartphone ausführen kann dass ich ihn auch in Papierform stellen kann dass ich den Status der Bearbeitung einsehen kann dass ich Beschwerdemöglichkeiten habe dass ich persönlichen Kontakt habe	that the application is understandable that my data is secure that little data is requested that it can be informal that I can do it from my smartphone that I can also submit it in paper form that I can see the status of the processing that I have the possibility to complain that I have personal contact

C.2 Questions Results

The following paragraphs show the partially referenced results of the questionnaire. The anonymised raw data can be found in the TU Wien archives.

Results of Question 1 Figure C.1 shows a bar graph of the results of Question 1.

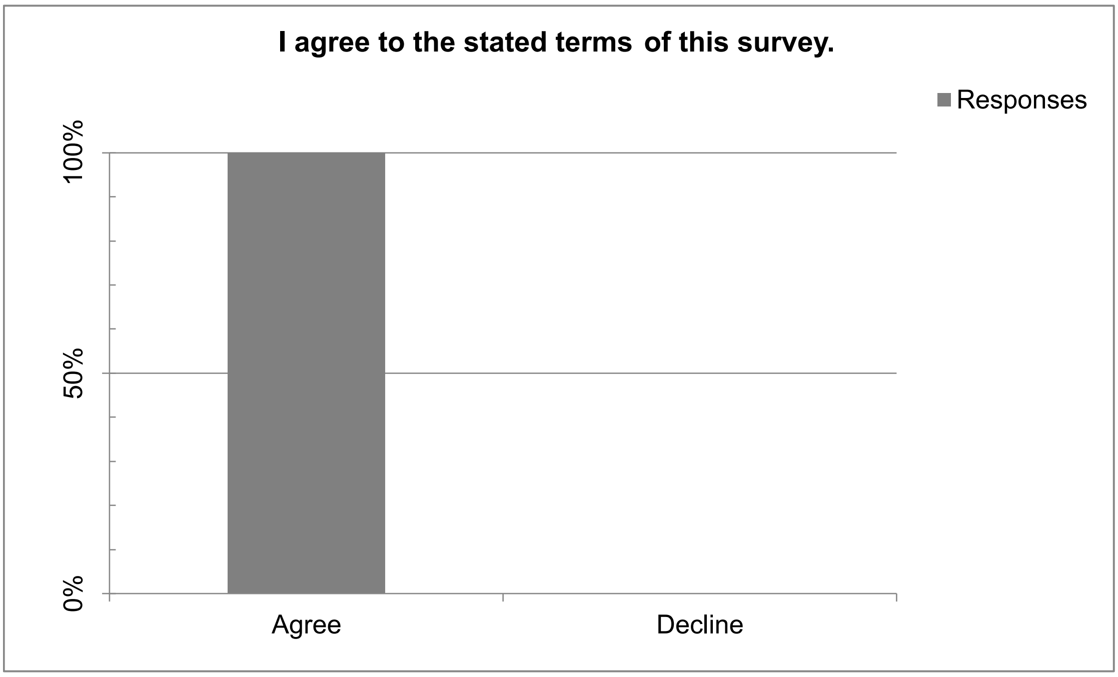


Figure C.1: Bar chart of survey question 1

Table C.2 shows the result values of Question 1.

Table C.2: Results of Question 1

I agree to the stated terms of this survey.

Answer Choices	**Responses**	
Agree	100.0%	99
Decline	0.0%	0
	Answered	99
	Skipped	0

Results of Question 2 Figure C.2 shows a bar graph of the results of Question 2.

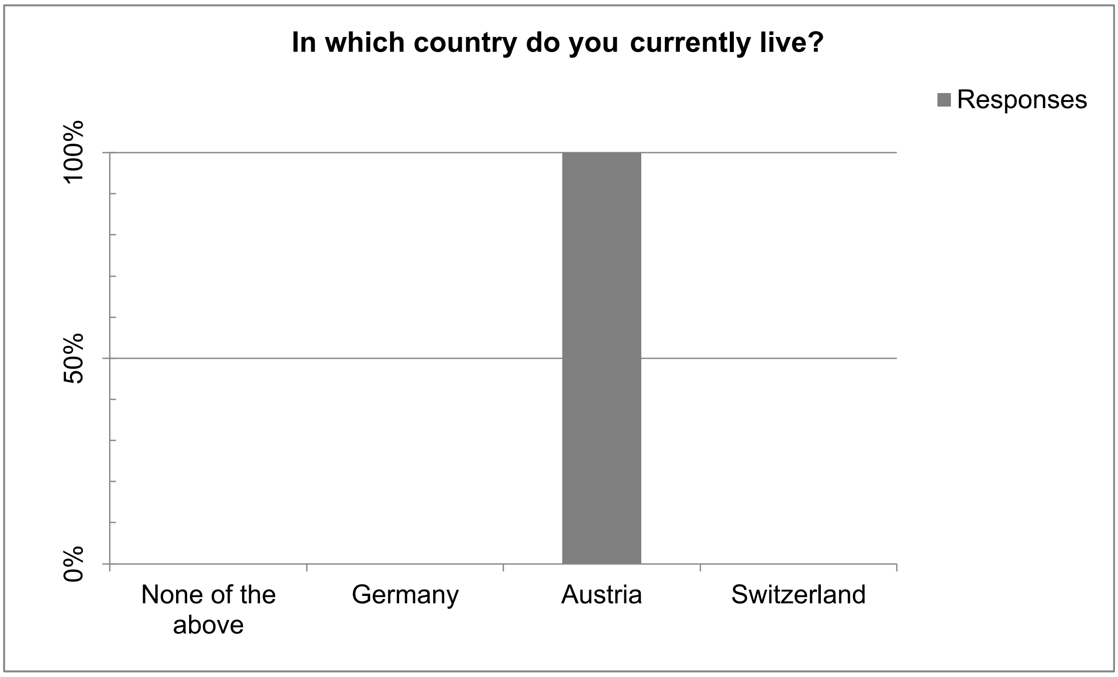


Figure C.2: Bar chart of survey question 2

Table C.3 shows the result values of Question 2.

Table C.3: Results of Question 2

In which country do you currently live?

Answer Choices	**Responses**	
None of the above	0.0%	0
Germany	0.0%	0
Austria	100.0%	99
Switzerland	0.0%	0
	Answered	99
	Skipped	0

Results of Question 3 Figure C.3 shows a bar graph of the results of Question 3.

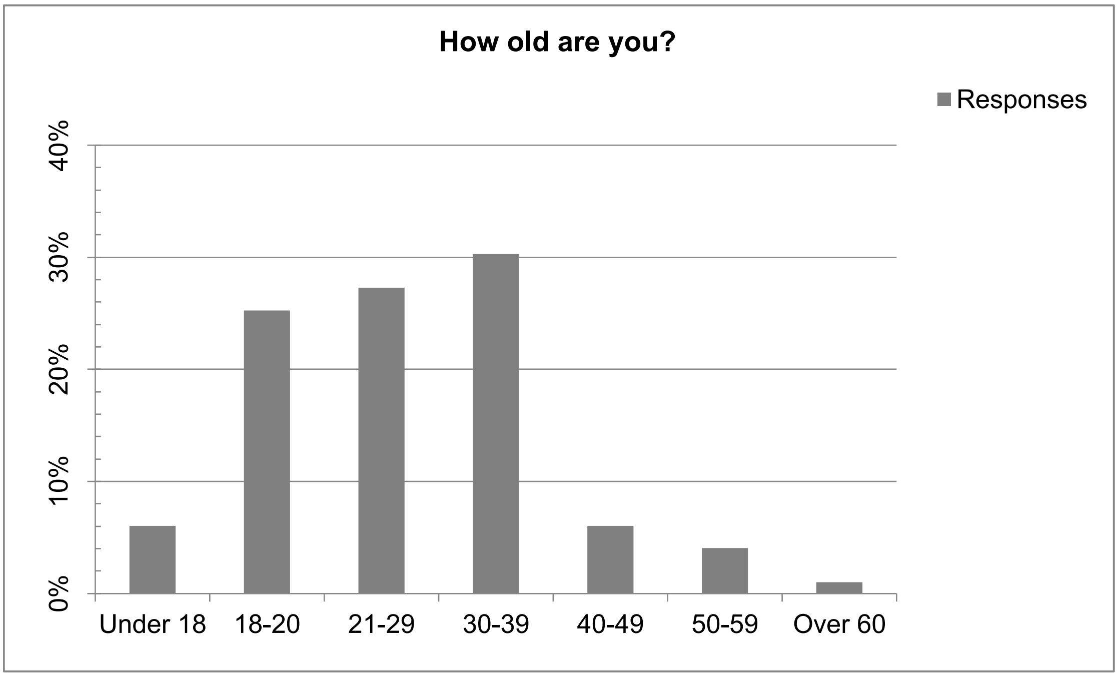


Figure C.3: Bar chart of survey question 3

Table C.4 shows the result values of Question 3.

Table C.4: Results of Question 3

How old are you?

Answer Choices	**Responses**	
Under 18	6.1%	6
18-20	25.3%	25
21-29	27.3%	27
30-39	30.3%	30
40-49	6.1%	6
50-59	4.0%	4
Over 60	1.0%	1
	Answered	99
	Skipped	0

Results of Question 4 Figure C.4 shows a bar graph of the results of Question 4.

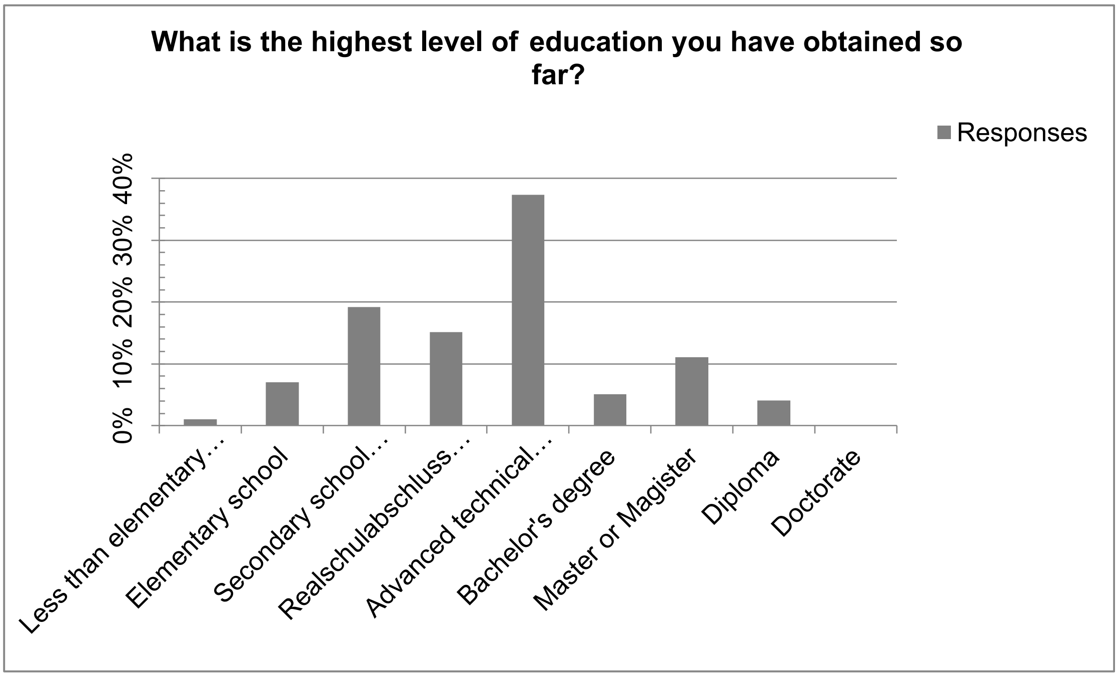


Figure C.4: Bar chart of survey question 4

Table C.5 shows the result values of Question 4.

Table C.5: Results of Question 4

What is the highest level of education you have obtained so far?

Answer Choices	**Responses**	
Less than elementary school	1.0%	1
Elementary school	7.1%	7
Secondary school diploma	19.2%	19
Realschulabschluss or Mittlere Reife (intermediate school leaving certificate)	15.2%	15
Advanced technical college certificate or high school diploma	37.4%	37
Bachelor's degree	5.1%	5
Master or Magister	11.1%	11
Diploma	4.0%	4
Doctorate	0.0%	0
	Answered	99
	Skipped	0

Results of Question 5 Figure C.5 shows a bar graph of the results of Question 5.

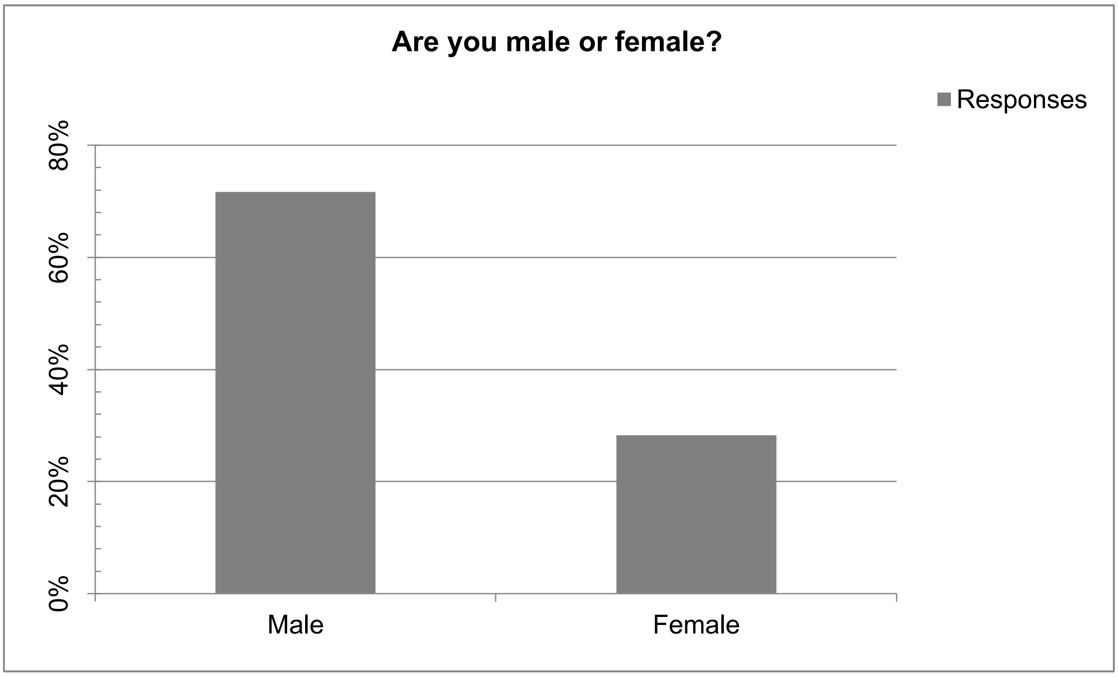


Figure C.5: Bar chart of survey question 5

Table C.6 shows the result values of Question 5.

Table C.6: Results of Question 5

Are you male or female?

Answer Choices	**Responses**	
Male	71.7%	71
Female	28.3%	28
	Answered	99
	Skipped	0

Results of Question 6 Figure C.6 shows a bar graph of the results of Question 6.

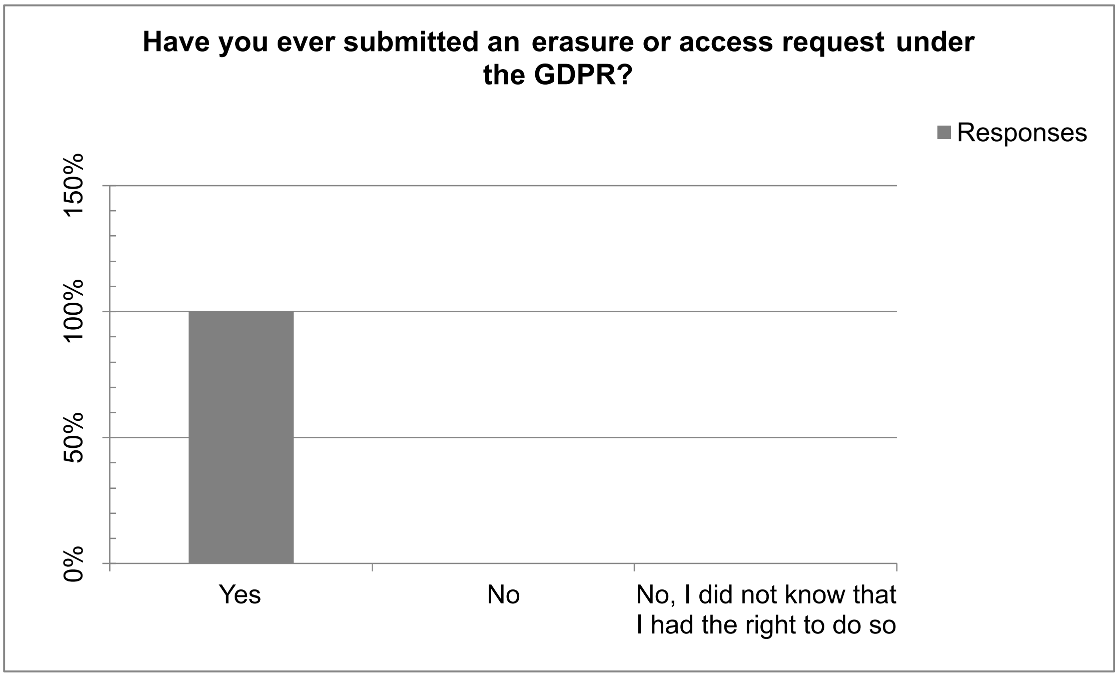


Figure C.6: Bar chart of survey question 6

Table C.7 shows the result values of Question 6.

Table C.7: Results of Question 6

Have you ever submitted an erasure or access request under the GDPR?

Answer Choices	**Responses**	
Yes	100.0%	99
No	0.0%	0
No, I did not know that I had the right to do so	0.0%	0
	Answered	99
	Skipped	0

Results of Question 7 Figure C.7 shows a bar graph of the results of Question 7.

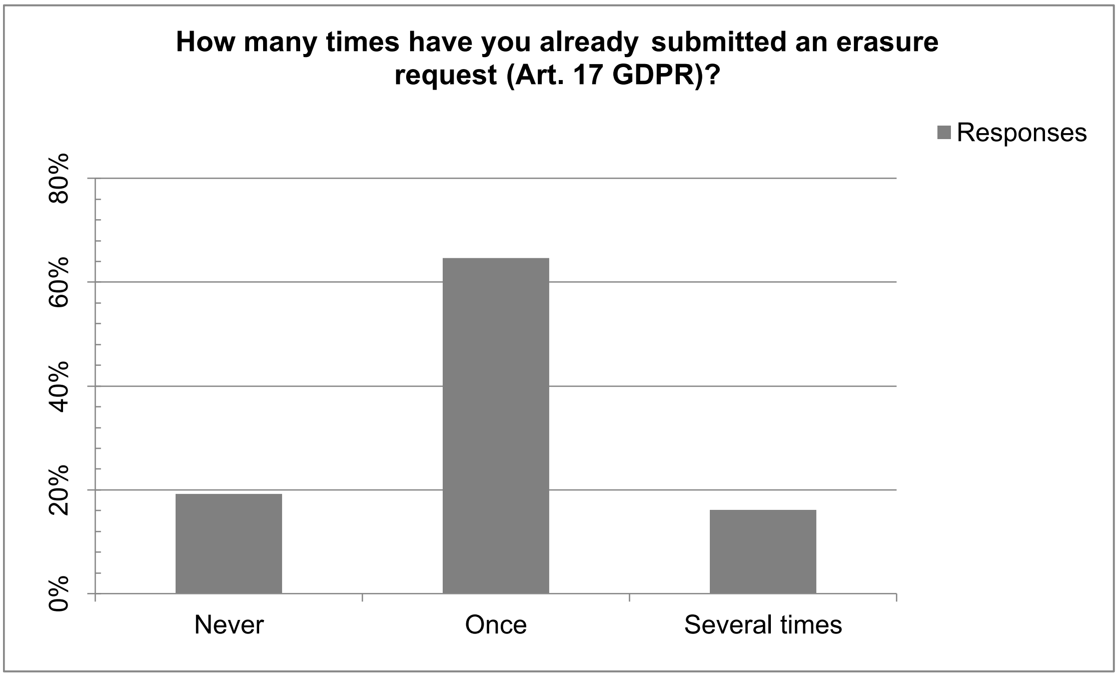


Figure C.7: Bar chart of survey question 7

Table C.8 shows the result values of Question 7.

Table C.8: Results of Question 7

How many times have you already submitted an erasure request (Art. 17 GDPR)?

Answer Choices	**Responses**	
Never	19.2%	19
Once	64.7%	64
Several times	16.2%	16
	Answered	99
	Skipped	0

Results of Question 8 Figure C.8 shows a bar graph of the results of Question 8.

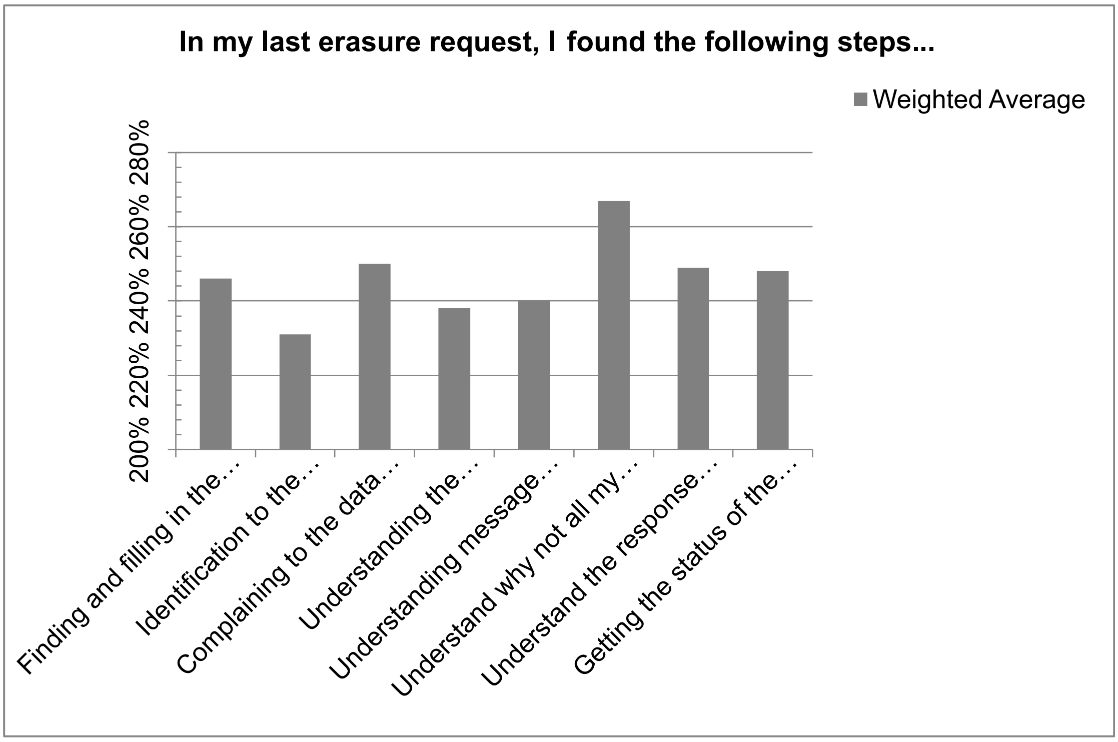


Figure C.8: Bar chart of survey question 8

Table C.9 shows the result values of Question 8.

Table C.9: Results of Question 8

In my last erasure request, I found the following steps...

	very simply		**simply**		**ok**		**difficult**		**very difficult**		**was not the case**		**Total**	**Weighted Average**
Finding and filling in the application form	21.3%	17	31.3%	25	31.3%	25	6.3%	5	7.5%	6	2.5%	2	80	2.46
Identification to the company (e.g. scan ID card)	25.0%	20	33.8%	27	26.3%	21	8.8%	7	3.8%	3	2.5%	2	80	2.31
Complaining to the data protection authority	13.8%	11	32.5%	26	28.7%	23	8.8%	7	3.8%	3	12.5%	10	80	2.5

Understanding the company's response	20.0%	16	37.5%	30	23.8%	19	8.8%	7	5.0%	4	5.0%	4	80	2.38
Understanding message about delay	16.3%	13	36.3%	29	31.3%	25	7.5%	6	2.5%	2	6.3%	5	80	2.4
Understand why not all my data was deleted	15.0%	12	27.5%	22	32.5%	26	13.8%	11	6.3%	5	5.0%	4	80	2.67
Understand the response of the complaint to the data protection authority	12.5%	10	36.3%	29	32.5%	26	11.3%	9	1.3%	1	6.3%	5	80	2.49
Getting the status of the request	16.3%	13	41.3%	33	25.0%	20	10.0%	8	6.3%	5	1.3%	1	80	2.48

													Answered	80
													Skipped	19

Results of Question 9 Figure C.9 shows a bar graph of the results of Question 9.

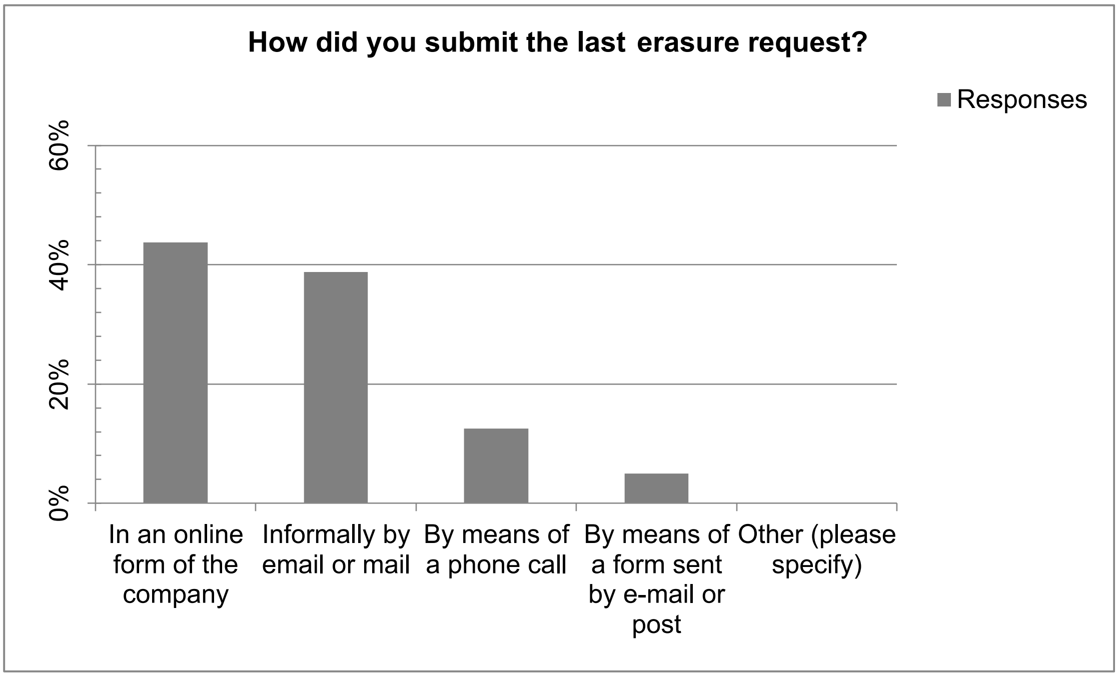


Figure C.9: Bar chart of survey question 9

Table C.10 shows the result values of Question 9.

Table C.10: Results of Question 9

How did you submit the last erasure request?

Answer Choices	**Responses**	
In an online form of the company	43.8%	35
Informally by email or mail	38.8%	31
By means of a phone call	12.5%	10
By means of a form sent by e-mail or post	5.0%	4
Other (please specify)	0.0%	0
	Answered	80
	Skipped	19

Results of Question 10 Figure C.10 shows a bar graph of the results of Question 10.

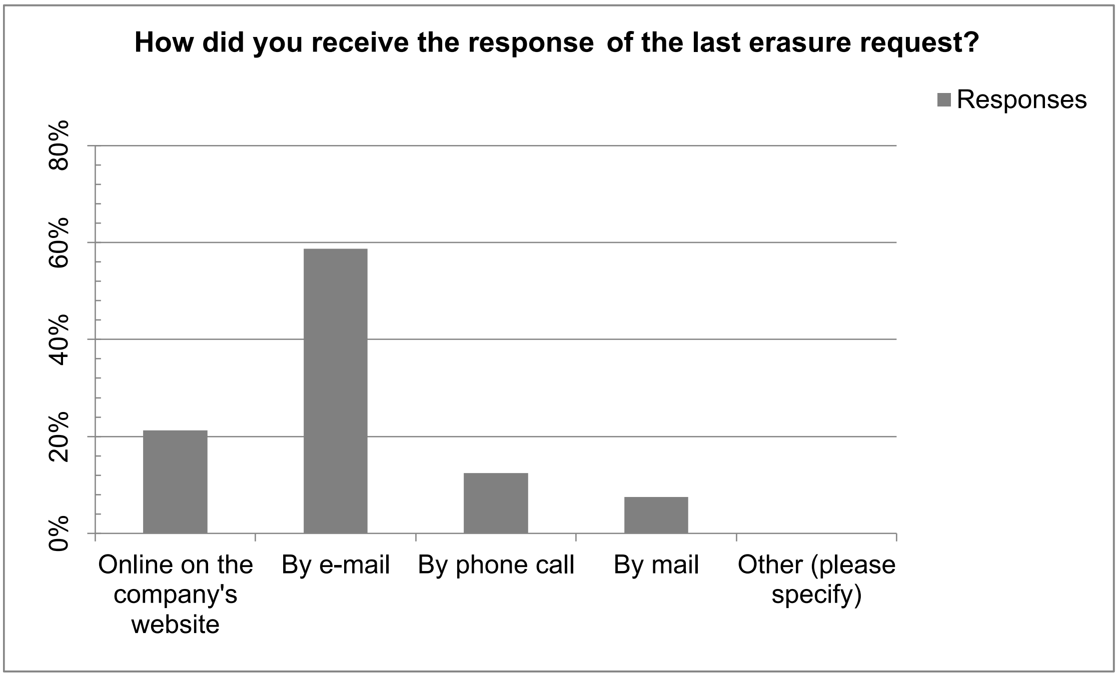


Figure C.10: Bar chart of survey question 10

Table C.11 shows the result values of Question 10.

Table C.11: Results of Question 10

How did you receive the response of the last erasure request?

Answer Choices	**Responses**	
Online on the company's website	21.3%	17
By e-mail	58.8%	47
By phone call	12.5%	10
By mail	7.5%	6
Other (please specify)	0.0%	0
	Answered	80
	Skipped	19

Results of Question 11 Figure C.11 shows a bar graph of the results of Question 11.

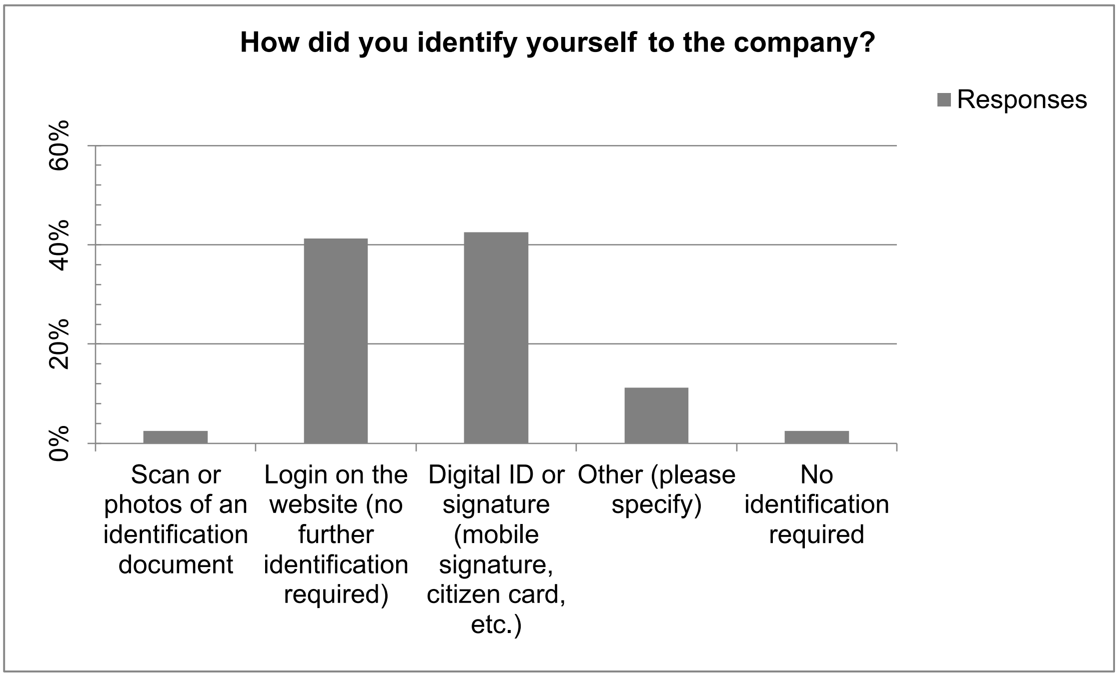


Figure C.11: Bar chart of survey question 11

Table C.12 shows the result values of Question 11.

Table C.12: Results of Question 11

How did you identify yourself to the company?

Answer Choices	Responses	
Scan or photos of an identification document	2.5%	2
Login on the website (no further identification required)	41.3%	33
Digital ID or signature (mobile signature, citizen card, etc.)	42.5%	34
Other (please specify)	11.3%	9
No identification required	2.5%	2
	Answered	80
	Skipped	19

Results of Question 12 Figure C.12 shows a bar graph of the results of Question 12.

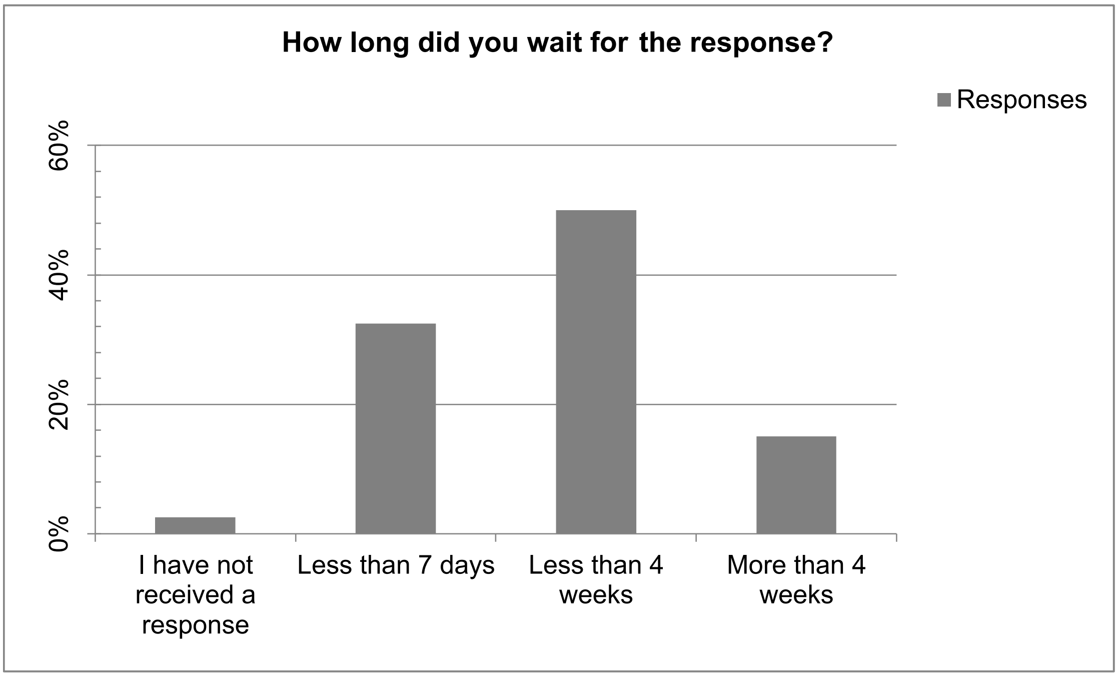


Figure C.12: Bar chart of survey question 12

Table C.13 shows the result values of Question 12.

Table C.13: Results of Question 12

How long did you wait for the response?

Answer Choices	**Responses**	
I have not received a response	2.5%	2
Less than 7 days	32.5%	26
Less than 4 weeks	50.0%	40
More than 4 weeks	15.0%	12
	Answered	80
	Skipped	19

Results of Question 13 Figure C.13 shows a bar graph of the results of Question 13.

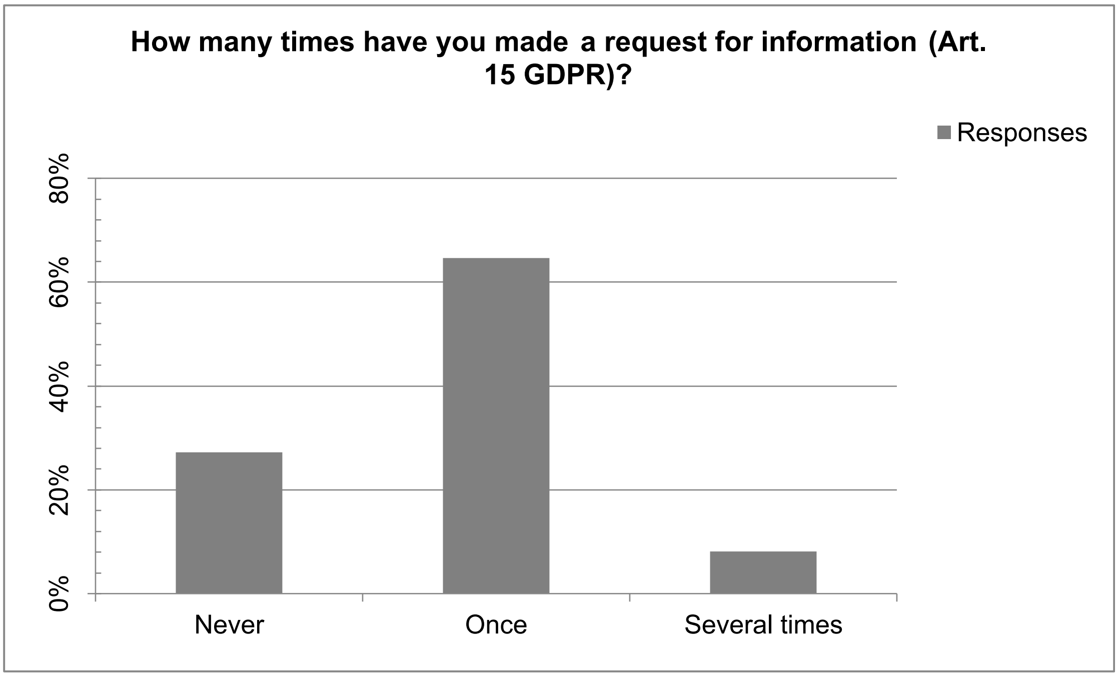


Figure C.13: Bar chart of survey question 13

Table C.14 shows the result values of Question 13.

Table C.14: Results of Question 13

How many times have you made a request for information (Art. 15 GDPR)?

Answer Choices	**Responses**	
Never	27.3%	27
Once	64.7%	64
Several times	8.1%	8
	Answered	99
	Skipped	0

Results of Question 14 Figure C.14 shows a bar graph of the results of Question 14.

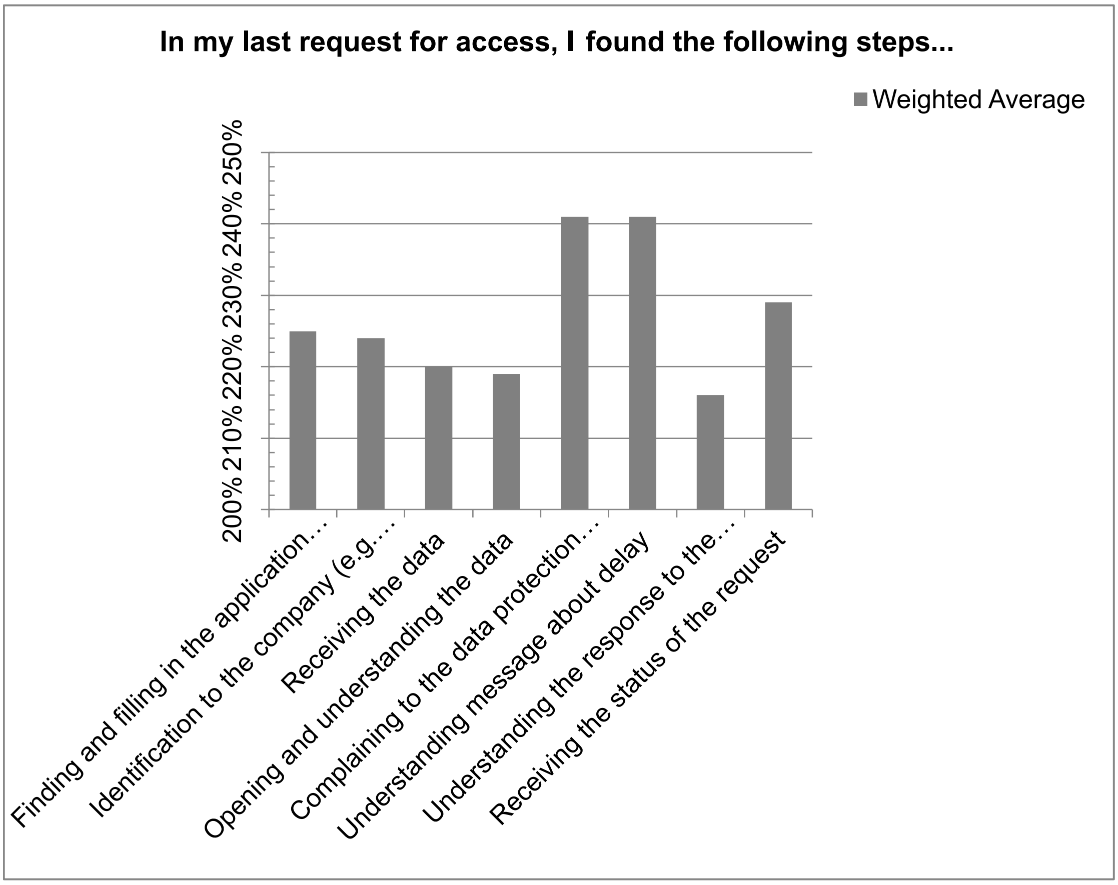


Figure C.14: Bar chart of survey question 14

Table C.15 shows the result values of Question 14.

Table C.15: Results of Question 14

In my last request for access, I found the following steps. . .	**very simply**		**simply**		**ok**		**difficult**		**very difficult**		**was not the case**		**Total**	**Weighted Average**
Finding and filling in the application form	20.8%	15	40.3%	29	30.6%	22	5.6%	4	1.4%	1	1.4%	1	72	2.25
Identification to the company (e.g. scan ID card)	20.8%	15	37.5%	27	27.8%	20	5.6%	4	1.4%	1	6.9%	5	72	2.24
Receiving the data	18.1%	13	52.8%	38	19.4%	14	6.9%	5	1.4%	1	1.4%	1	72	2.2
Opening and understanding the data	22.2%	16	47.2%	34	12.5%	9	13.9%	10	0.0%	0	4.2%	3	72	2.19

Complaining to the data protection authority	11.1%	8	34.7%	25	38.9%	28	4.2%	3	0.0%	0	11.1%	8	72	2.41
Understanding message about delay	13.9%	10	38.9%	28	29.2%	21	6.9%	5	2.8%	2	8.3%	6	72	2.41
Understanding the response to the complaint to the data protection authority	22.2%	16	38.9%	28	19.4%	14	4.2%	3	2.8%	2	12.5%	9	72	2.16
Receiving the status of the request	18.1%	13	41.7%	30	29.2%	21	4.2%	3	2.8%	2	4.2%	3	72	2.29
													Answered	72
													Skipped	27

Results of Question 15 Figure C.15 shows a bar graph of the results of Question 15.

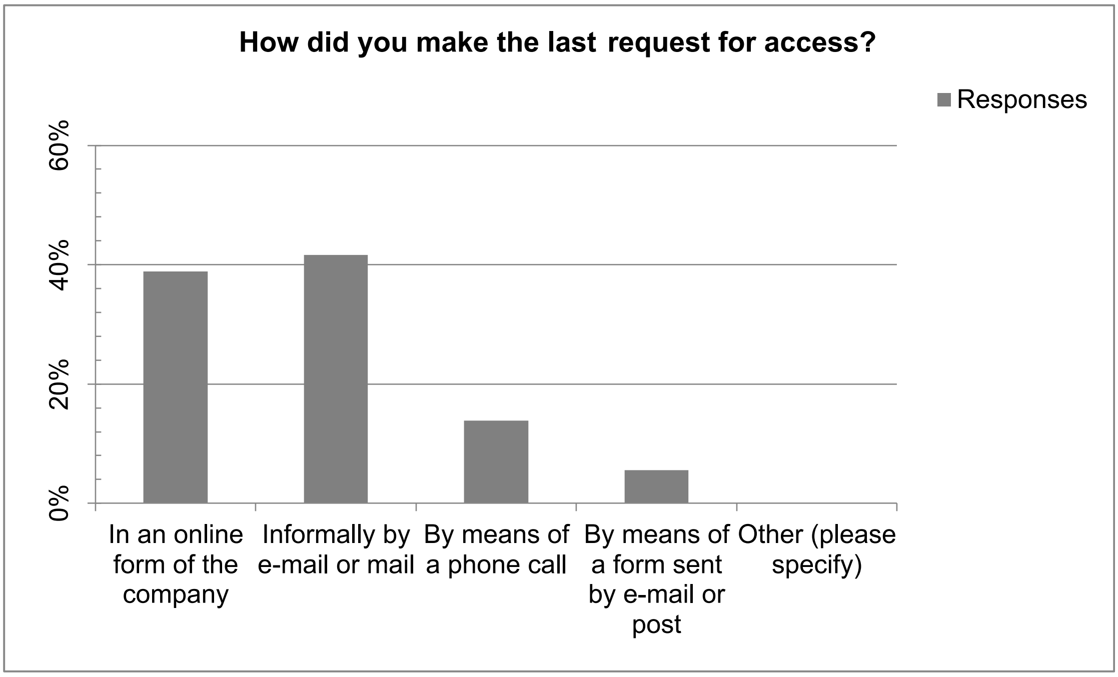


Figure C.15: Bar chart of survey question 15

Table C.16 shows the result values of Question 15.

Table C.16: Results of Question 15

How did you make the last request for access?

Answer Choices	Responses	
In an online form of the company	38.9%	28
Informally by e-mail or mail	41.7%	30
By means of a phone call	13.9%	10
By means of a form sent by e-mail or post	5.6%	4
Other (please specify)	0.0%	0
	Answered	72
	Skipped	27

Results of Question 16 Figure C.16 shows a bar graph of the results of Question 16.

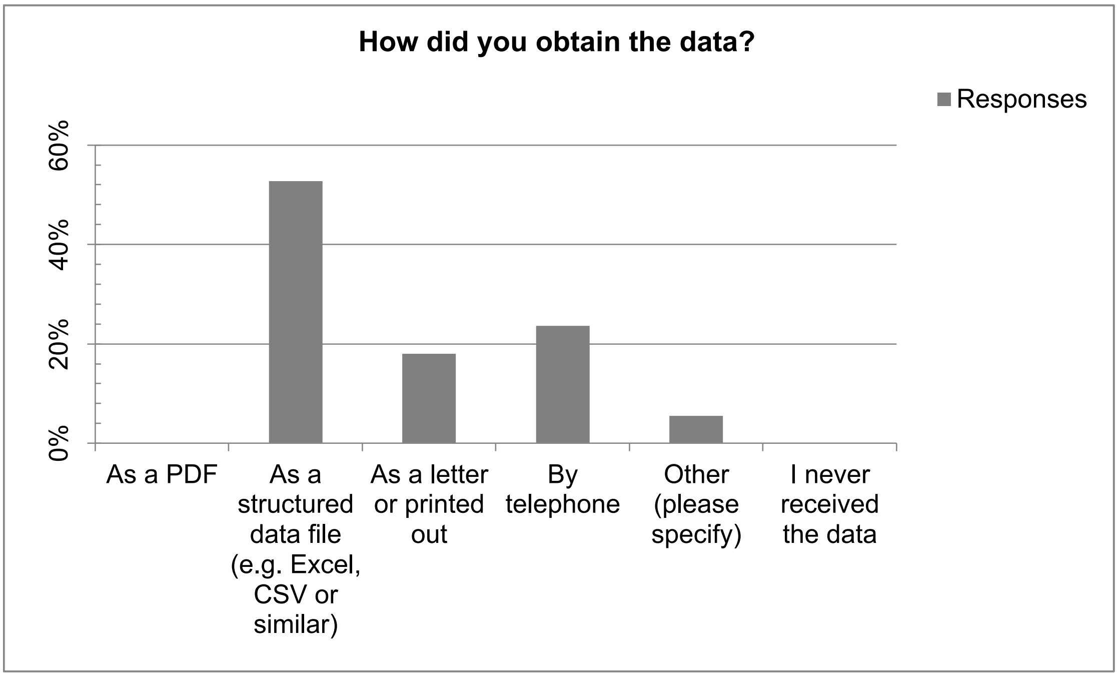


Figure C.16: Bar chart of survey question 16

Table C.17 shows the result values of Question 16.

Table C.17: Results of Question 16

How did you obtain the data?

Answer Choices	**Responses**	
As a PDF	0.0%	0
As a structured data file (e.g. Excel, CSV or similar)	52.8%	38
As a letter or printed out	18.1%	13
By telephone	23.6%	17
Other (please specify)	5.6%	4
I never received the data	0.0%	0
	Answered	72
	Skipped	27

Results of Question 17 Figure C.17 shows a bar graph of the results of Question 17.

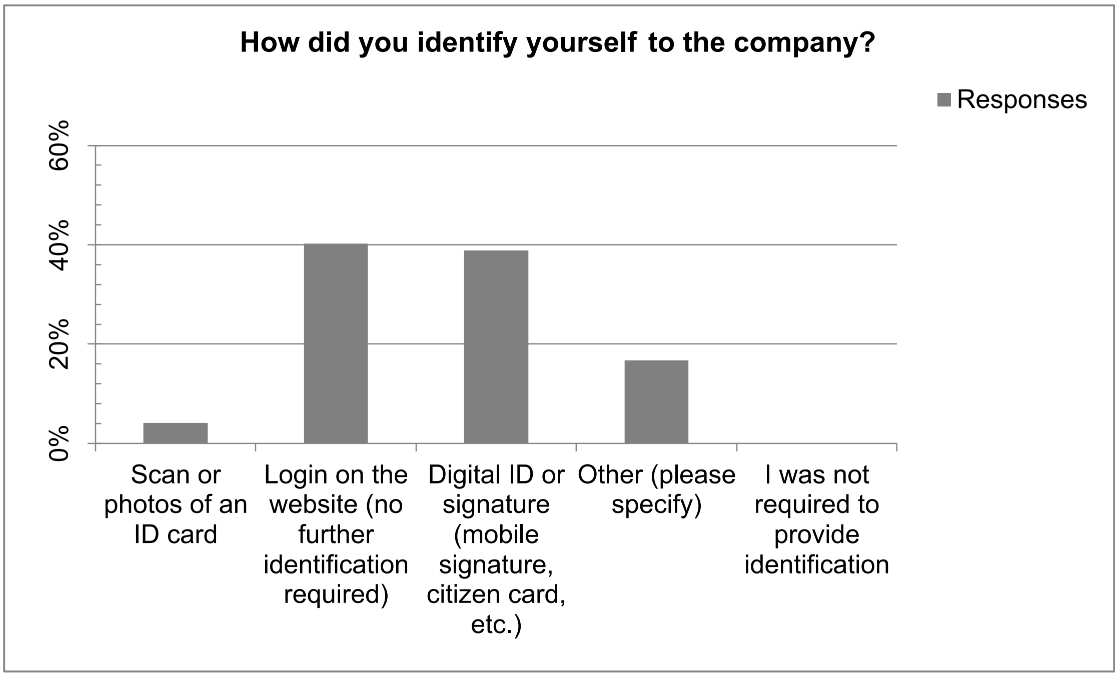


Figure C.17: Bar chart of survey question 17

Table C.18 shows the result values of Question 17.

Table C.18: Results of Question 17

How did you identify yourself to the company?

Answer Choices	**Responses**	
Scan or photos of an ID card	4.2%	3
Login on the website (no further identification required)	40.3%	29
Digital ID or signature (mobile signature, citizen card, etc.)	38.9%	28
Other (please specify)	16.7%	12
I was not required to provide identification	0.0%	0
	Answered	72
	Skipped	27

Results of Question 18 Figure C.18 shows a bar graph of the results of Question 18.

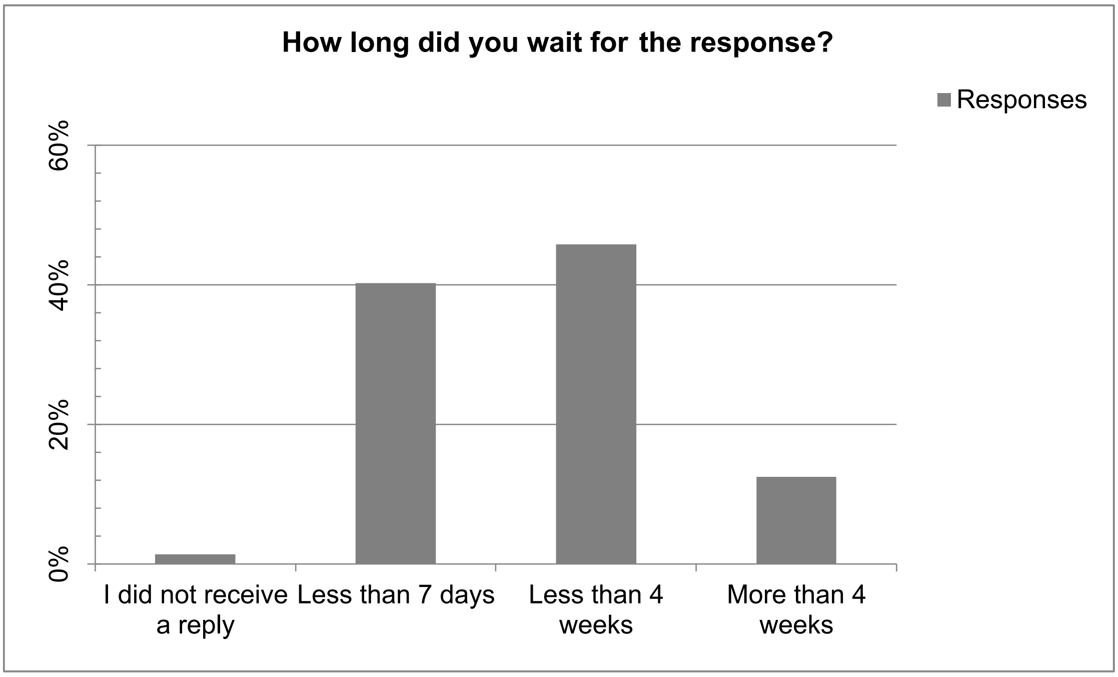


Figure C.18: Bar chart of survey question 18

Table C.19 shows the result values of Question 18.

Table C.19: Results of Question 18

How long did you wait for the response?

Answer Choices	**Responses**	
I did not receive a reply	1.4%	1
Less than 7 days	40.3%	29
Less than 4 weeks	45.8%	33
More than 4 weeks	12.5%	9
	Answered	72
	Skipped	27

Results of Question 19 Figure C.19 shows a bar graph of the results of Question 19.

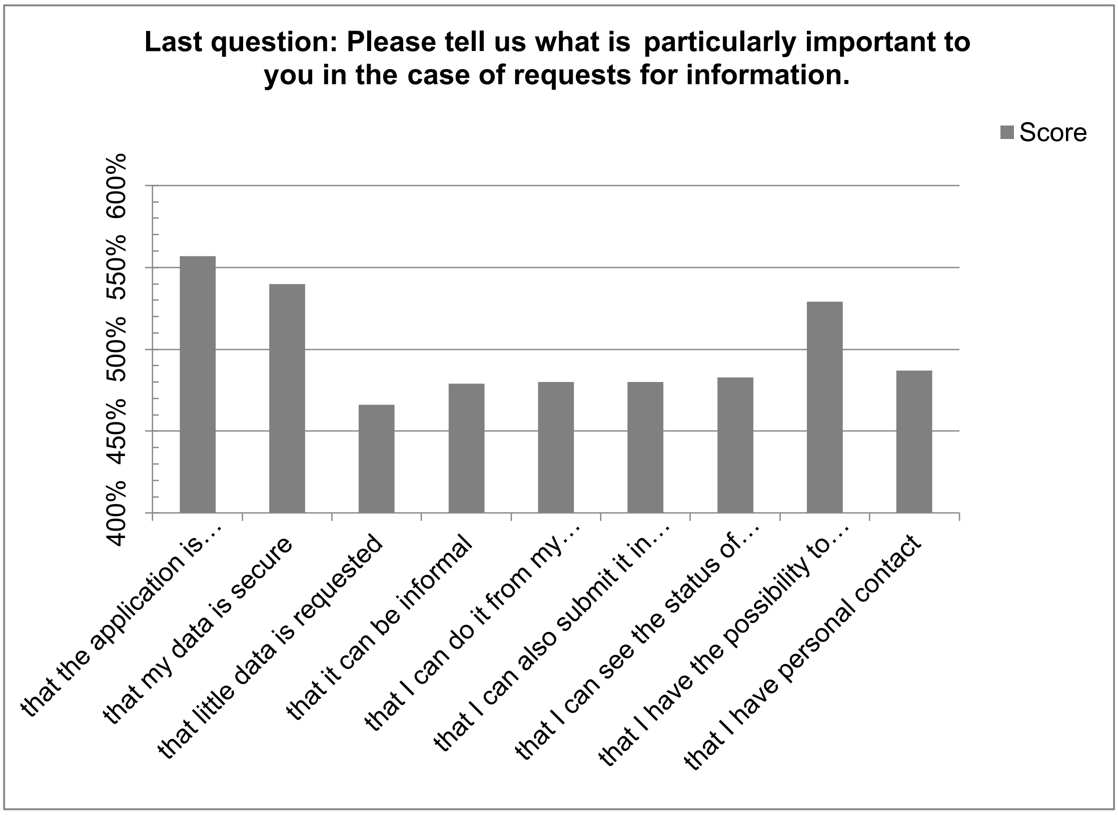


Figure C.19: Bar chart of survey question 19

Table C.20 shows the result values of Question 19.

Table C.20: Results of Question 19

Last question: Please tell us what is particularly important to you in the case of requests for information.

	1		**2**		**3**		**4**		**5**		**6**		**7**		**8**		**9**		**Total**	**Score**
that the application is understandable	13.1%	13	14.1%	14	15.2%	15	12.1%	12	9.1%	9	14.1%	14	9.1%	9	4.0%	4	9.1%	9	99	5.57
that my data is secure	21.2%	21	8.1%	8	8.1%	8	10.1%	10	12.1%	12	12.1%	12	8.1%	8	14.1%	14	6.1%	6	99	5.4
that little data is requested	7.1%	7	8.1%	8	9.1%	9	12.1%	12	19.2%	19	9.1%	9	12.1%	12	9.1%	9	14.1%	14	99	4.66
that it can be informal	12.1%	12	9.1%	9	9.1%	9	11.1%	11	10.1%	10	9.1%	9	13.1%	13	14.1%	14	12.1%	12	99	4.79
that I can do it from my smartphone	13.1%	13	10.1%	10	9.1%	9	11.1%	11	9.1%	9	10.1%	10	8.1%	8	11.1%	11	18.2%	18	99	4.8

that I can also submit it in paper form	6.1%	6	11.1%	11	15.2%	15	9.1%	9	13.1%	13	8.1%	8	15.2%	15	10.1%	10	12.1%	12	99	4.8
that I can see the status of the processing	8.1%	8	11.1%	11	7.1%	7	13.1%	13	14.1%	14	14.1%	14	10.1%	10	13.1%	13	9.1%	9	99	4.83
that I have the possibility to complain	12.1%	12	11.1%	11	18.2%	18	8.1%	8	10.1%	10	11.1%	11	9.1%	9	13.1%	13	7.1%	7	99	5.29
that I have personal contact	7.1%	7	17.2%	17	9.1%	9	13.1%	13	3.0%	3	12.1%	12	15.2%	15	11.1%	11	12.1%	12	99	4.87
																			Answered	99
																			Skipped	0

APPENDIX D

Appendix: Experts Delphi Rounds Regarding Implementation Criteria

D.1 Design of Questionnaire Round 1: Identification of Criteria

The following questionnaire contains an introduction to two GDPR processes and questions regarding these processes. The aim is to find criteria to evaluate software trying to help the particular process steps. You may end the questionnaire any time; your answers will be used for scientific research and will be published in an anonymized way.

D.1.1 Questions regarding Request for Erasure

Introduction A request to be 'forgotten' or 'request for erasure' is a request a natural person (called data subject) can file against a party that is processing its personal data and decides the purposes and means of the processing (called a controller) of the natural person according to the GDPR (Art. 17). The controller and processor then need to delete the data (with some exceptions). The following diagram shows the process of filing such a request against a controller that has one subcontractor (processor) who also processes the data on behalf of the controller.

Optional annotation information: The horizontal 'lanes' show the person or role executing the activity (each block). The white circle is the start and the bold white circle the end. The arrows show the 'flow' of execution whereas the 'x' symbol in the diamond shows an 'exclusive' gateway, meaning that only one of the following activities are executed. The clock symbol shows a waiting time and the 'lightning' an error. The filled envelop marks an activity that sends a message – the white envelop receives a message.

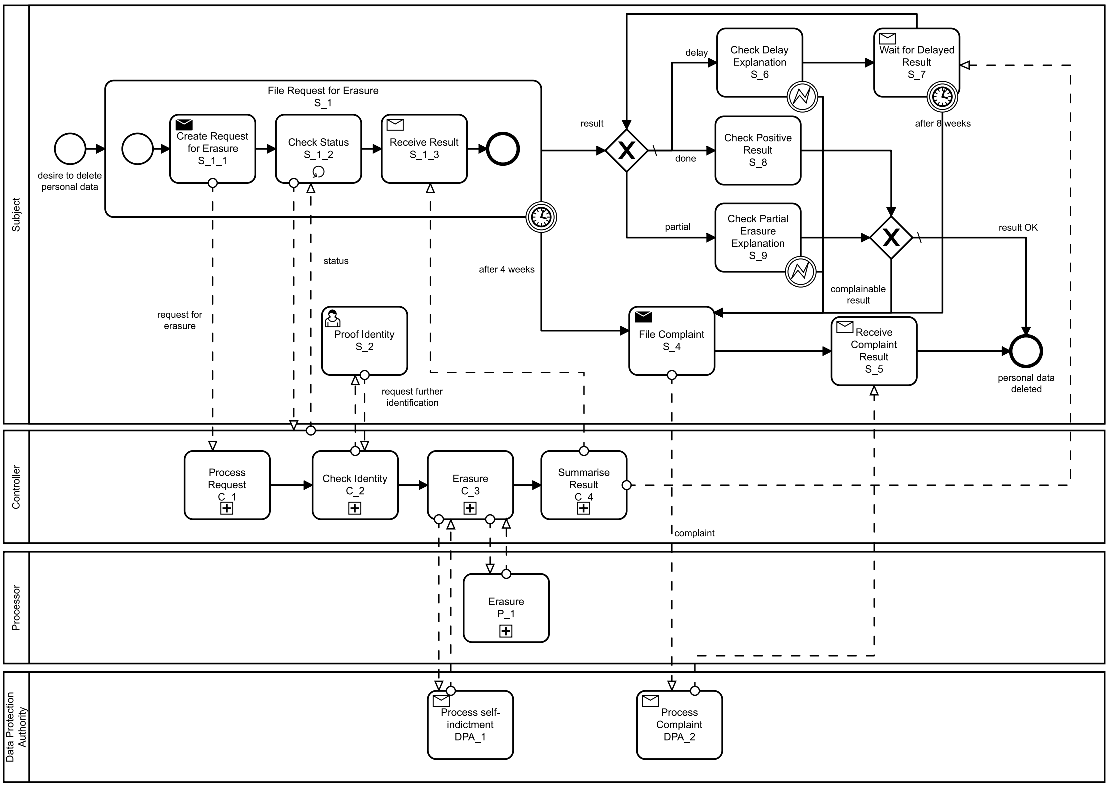


Figure D.1: Process of a request for erasure

1. Do you have any questions or comments regarding the process? Next, we want to find evaluation criteria for the activities of the data subject. An evaluation criterion is any factor that decides the quality of the activity for the data subject. For example, a car could have the following evaluation criteria: Speed, Safety, Fuel consumption, etc. 2. Can you argue any criteria that needs to be fulfilled by a software supporting each of the 10 subject's activities?

Table D.1: Questionnaire Erasure

Activity	Description	Input
Create Request for Erasure S_1_1	The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.	[Text Input]

Check Status S_1_2	The data subject checks whether the request has been done.	[Text Input]
Receive Result S_1_3	The data subject receives a response from the data controller.	[Text Input]
Proof Identity S_2	The data subject proofs its identity to the data controller on request.	[Text Input]
Check Delay Explanation S_6	The data subject receives and checks a notification of a delayed result with an explanation to the data subject.	[Text Input]
Wait for Delayed Result S_7	If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 8 weeks of waiting, the data subject may file a complaint.	[Text Input]
Check Positive Result S_8	The data subject receives and checks a positive result (data erased).	[Text Input]
Check Partial Erasure Explanation S_9	The data subject receives and checks a result stating that only parts of its data have been deleted.	[Text Input]
File Complaint S_4	The data subject files a complaint at the Data Protection Authority.	[Text Input]
Receive Complaint Result S_5	The data subject receives the result of the complaint.	[Text Input]

D.1.2 Questions regarding Request for Access

Introduction A ‘request to access’ is a request a data subject can file against a controller. The controller must then deliver information regarding the processing activities, the rights of the data subject and the data itself. The following diagram shows the process of filing such a request against a controller that has one processor.

1. Do you have any questions or comments regarding the process? [Text Input] Next, we again want to find evaluation criteria for the activities of the data subject. 2. Can you argue any criteria that needs to be fulfilled by a software supporting each of the 10 subject’s activities? (you may also use the same criteria as above if they fit)

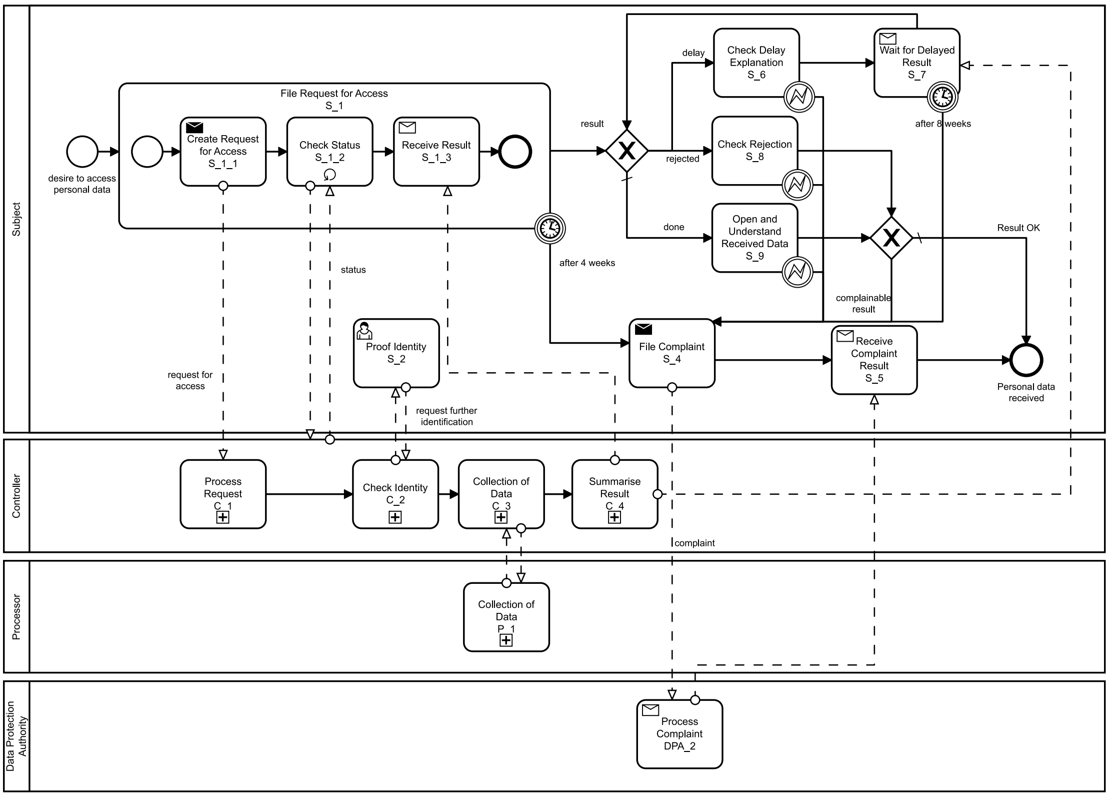


Figure D.2: Process of a request for access

Questionnaire Access

Activity	Description	Input
Create Request for Access S_1_1	The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.	[Text Input]
Check Status S_1_2	The data subject checks whether the request has been done.	[Text Input]
Receive Result S_1_3	The data subject receives a response from the data controller.	[Text Input]
Proof Identity S_2	The data subject proofs its identity to the data controller on request.	[Text Input]

Check Delay Explanation S_6	The data subject receives and checks a result stating that only parts of its data have been deleted.	[Text Input]
Wait for Delayed Result S_7	If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 8 weeks of waiting, the data subject may file a complaint.	[Text Input]
Check Rejection S_8	If the response was a rejection, the data subject checks the reason for the rejection.	[Text Input]
Open and Understand Received Data S_9	After receiving the data files, the data subject tries to open and understand the data.	[Text Input]
File Complaint S_4	The data subject files a complaint at the Data Protection Authority.	[Text Input]
Receive Complaint Result S_5	The data subject receives the result of the complaint.	[Text Input]

D.2 Results of Questionnaire Round 1: Identification of Criteria

The following section shows the unfiltered freely given responses to the questions to find criteria regarding the process activities.

D.2.1 Results regarding Request for Erasure

Results of Question 1: Do you have any questions or comments regarding the process?

- One minor, possibly not so relevant, remark: After the subject has received the result of the complaint from the DPA (*S_5*) it is concluded that in the immediate next step their data is erased. Wouldn't there be an involvement of the controller necessary in order to execute the DPA's complaint, which then would lead to said data to be erased?
- Dependent on the concrete use case an additional process between *C_2* (ID Check) and *C_3* (Erasure) would have to be added: Check, if the erasure is possible (e.g. in case contractual obligations prevent this - see Art 17 para 3 GDPR). I assume process *DPA_1* is for the case that the personal data was unlawfully processed.
- The process appears to be somewhat convoluted.
- Not that familiar with the notation but 'Summarize results' is received twice. Once as a Result and once as a delayed result.
- Not easy to understand at first glance, you have to get used to it.

Results of Question 2: Can you argue any criteria that needs to be fulfilled by a software supporting each of the 10 subject's activities? (Please use a dot '.' where you would use a new line or paragraph). Create Request for Erasure *S_1_1* The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.

- User friendly selection which data shall be covered by the request for erasure
- Easy access (possibly multiple channels?) for the subjects to file their requests. Easy handling, e.g. generated form, which can be filed within a few clicks
- usability. completeness of information in the form. time needed to fill the form / percentage of manual input needed. confirmation sent after successful request
- fault tolerance
- Simplicity and safety: Online, phone, website, written letter, e-Mail.
- The controller's contact info must be available to the subject.
- Convenience. User-friendliness.
- Time required to find way of communication. Accessability of communication also for disabled users.
- Time it takes to discover the information. Providing a template, Providing a step by step explanation of the process
- Usability

Results of Question 3: Check Status S_1_2 The data subject checks whether the request has been done.

- Frequent Automatic updates
- Timley generated confirmation message.
- availability. time since request was sent
- predictability
- Simplicity: Simple unique link, that can be checked, displaying the status; SMS and/or E-Mail update.
- There must be a way for the subject to gather this information.
- Understandability.
- Status Updates shall be near real time with an audit trail.
- Time to log into the system
- Usability

Results of Question 4: Receive Result S_1_3 The data subject receives a response from the data controller.

- Receive Notification
- Timley response (within the legally required response time, otherwise state reasons for delayed response). Include all legal relevant information (e.g. complete erasure, partial erasure, reasons for denial of request) in "clear and plain language" (Art 12 para 1)
- push notification. availability at a later time

- consistency
- Simplicity: Same or similar medium of communication as when receiving the request for erasure: Written letter, E-Mail, SMS.
- The subject's contact info must be available to the controller.
- Understandability.
- Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.)
- How many types of notification are supported
- Usability, Privacy

Results of Question 4: Proof Identity S_2 The data subject proofs its identity to the data controller on request.

- Easy verification process
- Identification via official identification document (driver's licence, identity card, passport), threshould concerning identity is "reasonable doubt"
- time needed for validation. document options for validation. online yes/no. mobile yes/no
- security
- Safety; Online via webcam and passport/id; digital identification (e.g. Buergerkarte), depending on use case (e.g. non-governmental, simple online services) by username and password.
- The subject has a proof of identity available. They are willing to share this with the controller. They have a way to send this information to the controller.
- User-friendliness.
- Software shall check if proof identiy quality is good enough for further processing
- Number of prerequisites the user needs to proof this.
- Trust

Results of Question 5: Check Delay Explanation S_6 The data subject receives and checks a notification of a delayed result with an explanation to the data subject.

- Explanation of rights upon expiry of the deadline
- Ensure to be within legally allowed response time
- clarness of explanation
- consistency
- Simplicity, safety, speed (immediate response as soon as a delay is obvious).
- There must be a way for the subject to gather this information.
- Confirmability.
- Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.). Also possibility to get in touch with the processor.
- Same as check status S_1_2
- Usability

Results of Question 6: Wait for Delayed Result S_7 If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 8 weeks of waiting, the data subject may file a complaint.

- Notification of expiry of deadline. Provide (link) form for complaint.
- Minor remark: to be precise, after two months (not 8 weeks) the data subject may file a complaint. Ensure to be within legally allowed response time.
- trigger after x weeks yes/no
- predictability
- Simplicity: Automated generation of a complaint (at this point the ID is clear, the issue should be, the responsible authority is clear).
- There must be a way for the subject to gather this information.
- Convenience. User-friendliness. Possibility to discharge anger.
- Complaint shall be automaticially filed if "okeyed" initally. I.e. "do you want to automatically file a complaint after 8 weeks of non-resolution? yes/no"
- (Shouldn't that happen automatically, e.g. require this process to be implemented at the provider side)
- Speed (8 weeks??)

Results of Question 7: Check Positive Result S_8 The data subject receives and checks a positive result (data erased).

- Options to export and save the result with history.
- Ensure to be within legally allowed response time.
- clearness of result and what was deleted
- accuracy
- Simplicity: Overview of the process and outcome.
- .
- Convenience. Clearity.
- Full Report with (process) audit rail signed by the data processor.
- Check if necessary information is included in the result.
- no comment

Results of Question 8: Check Partial Erasure Explanation S_9 The data subject receives and checks a result stating that only parts of its data have been deleted.

- Provide potential reasons for the result. Explain options moving forward.
- Explicitly state which data have been delated and which not. Provide clear and concise reasons for only partial erasure.
- clearness of result and what was deleted
- traceability
- Simplicity: Overview of the process and outcome, possible remedies plus link/further info (e.g. automated complaint generation).
- .

- Convenience. Clearity. Understandability. Easy possibility to file a complaint to the Authority.
- Ability to request deletion of the remaining data and possiblity to have a report on what data has been deleted and what has not been deleted. Also explanation on a per data item basis why no deletion was done yet or why it can't be done.
- Does the explanation contain technical terms?
- no comment

Results of Question 10: File Complaint S_4 The data subject files a complaint at the Data Protection Authority.

- Instruction on formal requirements. Verify submission.
- Easy submittal: understandable language, form should be possible to file within a few clicks
- template for filing complaint?
- accessibility
- Simplicity and safety: Automated generation of complaint as far as possible (see S_7).
- .
- Convenience. Clearity. Understandability.
- Should be automated with a full collection of all steps that happend so far (audit trail)
- Is there template to guide through the process.
- Trust

Results of Question 11: Receive Complaint Result S_5 The data subject receives the result of the complaint.

- Notification. Provide options moving forward.
- Provide data subject with further if the complaint failed (e.g. appeal against decision of DPA)
- clearness of result and what was (not) deleted
- comprehensibility
- Simplicity: Overview of the process and outcome; information on further remedies/steps possible.
- .
- Convenience. Clearity. Understandability.
- n/a
- Does it contain instructions to what to do next?
- Trust and Privacy

- **D.2.2 Results regarding Request for Access**

Results of Question 12: Do you have any questions or comments regarding the process?

- One minor remark: GDPR uses "months" instead of "weeks" in this regard; also applicable for the first diagram
- Only major difference seems to be "self-indictment". I could not find by Googling what is meant by this. I feel criteria are very similar to the process for deletion
- Dependent on the concrete use case an additional process between C_2 (ID Check) and C_3 (Erasure) would have to be added: Check, if the erasure is possible (e.g. in case contractual obligations prevent this - see Art 17 para 3 GDPR). I assume process DPA_1 is for the case that the personal data was unlawfully processed.
- no comment

Results of Question 13: Can you argue any criteria that needs to be fulfilled by a software supporting each of the 10 subject's activities? (you may also use the same criteria as above if they fit) Create Request for Access S_1_1 The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.

- User friendly selection which data shall be accessed
- Easy access, it should be possible to file request within a few clicks
- see proces for deletion
- fault tolerance for user input
- Simplicity: online, phone, website, e-mail, written letter
- The software must provide a way to enable this communication.
- User-friendliness. Easiness. Multilingualism.
- Time required to find way of communication. Accessability of communication also for disabled users.
- -//-
- Speed, Trust

Results of Question 14: Check Status S_1_2 The data subject checks whether the request has been done.

- Frequent automatic updates
- timley response (generated message), that request will be dealt with accordingly (maybe provide the legally allowed response time for the result)
- see proces for deletion
- predictability
- Simplicity: E.g. unique link displaying the current status, updates by e-mail or SMS
- The software must provide a way to enable this check.
- User-friendliness. Clearity.

- Status Updates shall be near real time with an audit trail.
- -//-
- Trust

Results of Question 15: Receive Result S_1_3 The data subject receives a response from the data controller.

- Receive notification
- As in all communication with the data subject, clear and easy understandable language. Ensure response within legally allowed time frame
- see proces for deletion
- consistency
- Simplicity: By the same or similar medium as the request for erasure was received (written letter, e-mail, SMS).
- The software must provide a way to enable this communication.
- User-friendliness. Clearity.
- Data Shall by encrypted with the data subjects key. Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.)
- -//-
- Speed, Usability

Results of Question 16: Proof Identity S_2 The data subject proofs its identity to the data controller on request.

- Easy verification process
- clear identification (ID card, passport, driver's licence), if the controller has reasonable doubts about the ID of the data subject. Assurance to data subject that their ID is only used for the identification process and deleted accordingly afterwards.
- see proces for deletion
- localized
- Safety and simplicity: Online via webcam and passport/ID; digital identification (e.g. Buergerkarte), in case of simple online services (non-governmental/-financial) via username and password.
- The subject must be willing to share this proof of identity. The software must provide a way proof the subject's identity.
- User-friendliness.
- Software shall check if proof identiy quality is good enough for further processing
- -//-
- Trust

Results of Question 17: Check Delay Explanation S_6 The data subject receives and checks a result stating that only parts of its data have been deleted.

- Provide explanation. Options moving forward.

- Clear and full and timley explanation of the reasons of the delay. (Minor remark: in this diagram it's rather about the access to data and not the deletion, right?)
- see proces for deletion
- comprehensible
- Simplicity: Overview of process and outcome, reasoning for the non-deletion; possible remedies.
- .
- User-friendliness. Understandability.
- Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.). Also possibility to get in touch with the processor.
- s/deleted/collected/ in the question, otherwise -//-
- Speed

Results of Question 18: Wait for Delayed Result S_7 If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 8 weeks of waiting, the data subject may file a complaint.

- Notification of expiry of deadline. Provide (link) form for complaint.
- Ensure to be within the legally required response time
- see proces for deletion
- predictability
- Simplicity and safety: Information after 8 weeks; auto-generation of a complaint (at this point the id of the person, the issue and the relevant authority should be known).
- .
- Easiness.
- Complaint shall be automaticially filed if "okeyed" initally. I.e. "do you want to automatically file a complaint after 8 weeks of non-resolution? yes/no"
- -//-
- Speed

Results of Question 19: Check Rejection S_8 If the response was a rejection, the data subject checks the reason for the rejection.

- Provide legal requirements for rejection.
- Provide clear and consice reasons for the rejection.
- see proces for deletion. clearness of explanation (what's the criterion? are there templates or pre-defined reasons? if you don't understand, can you ask?)
- comprehensible
- Simplicity and safety: Information on reasons, possible remedies and auto-generation of complaint (see S_7).
- .
- Understandability.
- Full Report with (process) audit rail signed by the data processor.

- -//-
- Speed, Trust

Results of Question 20: Open and Understand Received Data S_9 After receiving the data files, the data subject tries to open and understand the data.

- Options to save and export data.
- Provide not only the data files, but also a understandable explanation for data subjects not so familiar with this topic.
- see proces for deletion
- traceability.compatibility
- -
- .
- User-friendliness. Understandability.
- Data shall have human readable filenames that reflect the content of each file. data shall be in an industry standard fileformat.
- -//-
- Usability

Results of Question 21: File Complaint S_4 The data subject files a complaint at the Data Protection Authority.

- Instruction on formal requirements.
- Possibility to file the complaint within a few clicks (form) –> Usability
- see proces for deletion
- fault tolerance for user input
- Simplicity and safety: Auto-generation of complaint (see S_7).
- .
- User-friendliness. Multilingualism.
- Should be automated with a full collection of all steps that happend so far (audit trail)
- -//-
- Trust

Results of Question 22: Receive Complaint Result S_5 The data subject receives the result of the complaint.

- Notification. Provide legal background an solitons moving forward.
- Informing about further possibilities, if the result of the compliant is not satisfactory (e.g. appeal against descision of DPA)
- see proces for deletion
- accuracy
- Simplicity and safety: Overview of process and outcome, reasoning; information on further remedies against the decision.

- .
- User-friendliness. Understandability.
- n/a
- -//-
- Trust, Usability

D.3 Analysis of Questionnaire Round 1: Identification of Criteria

D.3.1 Analysis Step 1: Unfiltered Representation

The results are structured in a mindmap.

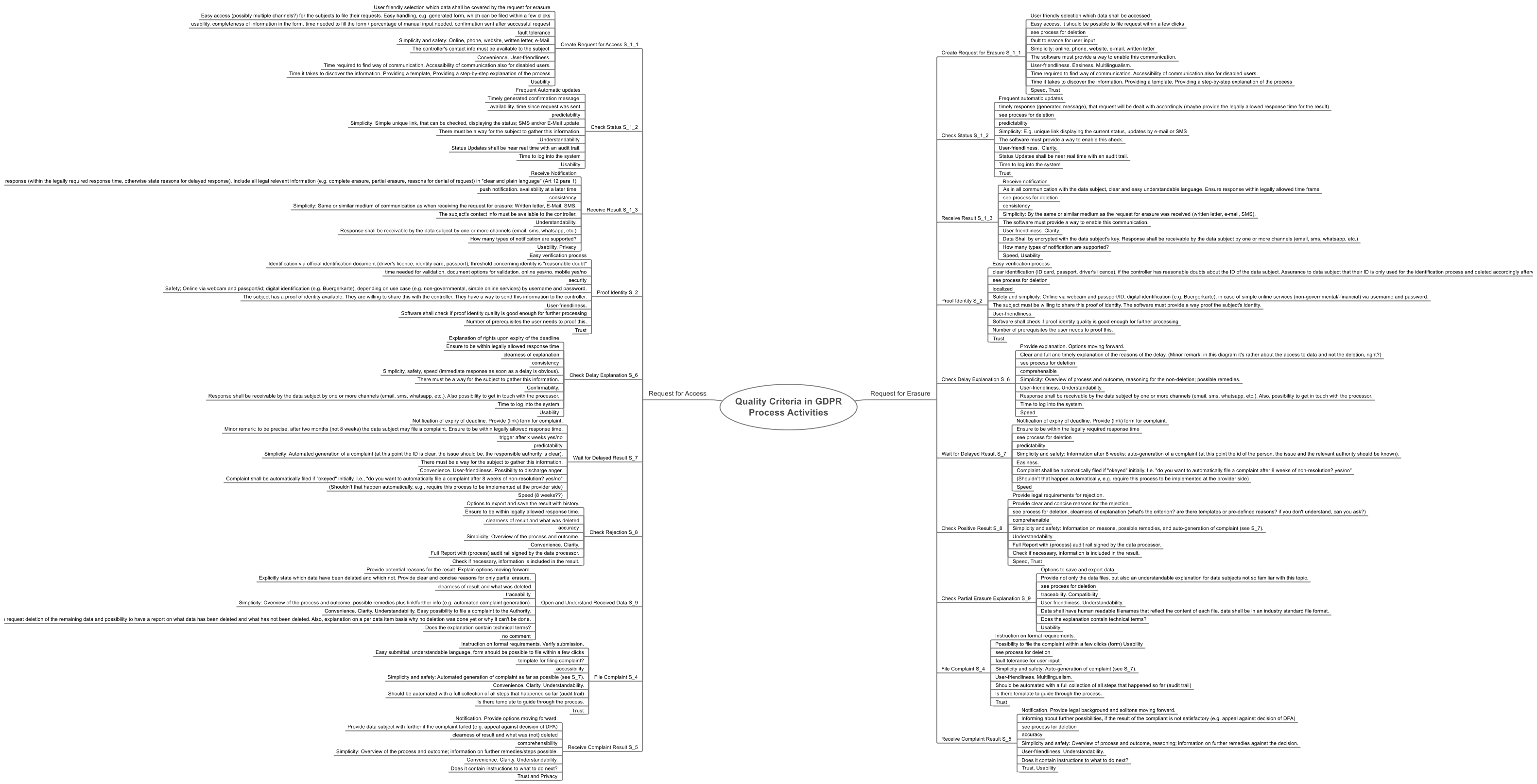

Quality Criteria in GDPR Process Activities
Request for Access
Create Request for Access S_1_1
User friendly selection which data shall be covered by the request for erasure
Easy access (possibly multiple channels?) for the subjects to file their requests. Easy handling, e.g. generated form, which can be filed within a few clicks
usability. completeness of information in the form. time needed to fill the form / percentage of manual input needed. confirmation sent after successful request
fault tolerance
Simplicity and safety: Online, phone, website, written letter, e-Mail.
The controller's contact info must be available to the subject.
Convenience. User-friendliness.
Time required to find way of communication. Accessibility of communication also for disabled users.
Time it takes to discover the information. Providing a template, Providing a step-by-step explanation of the process
Usability
Check Status S_1_2
Frequent Automatic updates
Timely generated confirmation message.
availability. time since request was sent
predictability
Simplicity: Simple unique link, that can be checked, displaying the status; SMS and/or E-Mail update.
There must be a way for the subject to gather this information.
Understandability.
Status Updates shall be near real time with an audit trail.
Time to log into the system
Usability
Receive Result S_1_3
Receive Notification
response (within the legally required response time, otherwise state reasons for delayed response). Include all legal relevant information (e.g. complete erasure, partial erasure, reasons for denial of request) in "clear and plain language" (Art 12 para 1)
push notification. availability at a later time
consistency
Simplicity: Same or similar medium of communication as when receiving the request for erasure: Written letter, E-Mail, SMS.
The subject's contact info must be available to the controller.
Understandability.
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.)
How many types of notification are supported?
Usability, Privacy
Proof Identity S_2
Easy verification process
Identification via official identification document (driver's licence, identity card, passport), threshold concerning identity is "reasonable doubt"
time needed for validation. document options for validation. online yes/no. mobile yes/no
security
Safety; Online via webcam and passport/id; digital identification (e.g. Buergerkarte), depending on use case (e.g. non-governmental, simple online services) by username and password.
The subject has a proof of identity available. They are willing to share this with the controller. They have a way to send this information to the controller.
User-friendliness.
Software shall check if proof identity quality is good enough for further processing
Number of prerequisites the user needs to proof this.
Trust
Check Delay Explanation S_6
Explanation of rights upon expiry of the deadline
Ensure to be within legally allowed response time
clearness of explanation
consistency
Simplicity, safety, speed (immediate response as soon as a delay is obvious).
There must be a way for the subject to gather this information.
Confirmability.
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.). Also possibility to get in touch with the processor.
Time to log into the system
Usability
Wait for Delayed Result S_7
Notification of expiry of deadline. Provide (link) form for complaint.
Minor remark: to be precise, after two months (not 8 weeks) the data subject may file a complaint. Ensure to be within legally allowed response time.
trigger after x weeks yes/no
predictability
Simplicity: Automated generation of a complaint (at this point the ID is clear, the issue should be, the responsible authority is clear).
There must be a way for the subject to gather this information.
Convenience. User-friendliness. Possibility to discharge anger.
Complaint shall be automatically filed if "okeyed" initially. I.e., "do you want to automatically file a complaint after 8 weeks of non-resolution? yes/no"
(Shouldn't that happen automatically, e.g., require this process to be implemented at the provider side)
Speed (8 weeks??)
Check Rejection S_8
Options to export and save the result with history.
Ensure to be within legally allowed response time.
clearness of result and what was deleted
accuracy
Simplicity: Overview of the process and outcome.
Convenience. Clarity.
Full Report with (process) audit rail signed by the data processor.
Check if necessary, information is included in the result.
Open and Understand Received Data S_9
Provide potential reasons for the result. Explain options moving forward.
Explicitly state which data have been delated and which not. Provide clear and concise reasons for only partial erasure.
clearness of result and what was deleted
traceability
Simplicity: Overview of the process and outcome, possible remedies plus link/further info (e.g. automated complaint generation).
Convenience. Clarity. Understandability. Easy possibility to file a complaint to the Authority.
request deletion of the remaining data and possibility to have a report on what data has been deleted and what has not been deleted. Also, explanation on a per data item basis why no deletion was done yet or why it can't be done.
Does the explanation contain technical terms?
no comment
File Complaint S_4
Instruction on formal requirements. Verify submission.
Easy submittal: understandable language, form should be possible to file within a few clicks
template for filing complaint?
accessibility
Simplicity and safety: Automated generation of complaint as far as possible (see S_7).
Convenience. Clarity. Understandability.
Should be automated with a full collection of all steps that happened so far (audit trail)
Is there template to guide through the process.
Trust
Receive Complaint Result S_5
Notification. Provide options moving forward.
Provide data subject with further if the complaint failed (e.g. appeal against decision of DPA)
clearness of result and what was (not) deleted
comprehensibility
Simplicity: Overview of the process and outcome; information on further remedies/steps possible.
Convenience. Clarity. Understandability.
Does it contain instructions to what to do next?
Trust and Privacy
Request for Erasure
Create Request for Erasure S_1_1
User friendly selection which data shall be accessed
Easy access, it should be possible to file request within a few clicks
see process for deletion
fault tolerance for user input
Simplicity: online, phone, website, e-mail, written letter
The software must provide a way to enable this communication.
User-friendliness. Easiness. Multilingualism.
Time required to find way of communication. Accessibility of communication also for disabled users.
Time it takes to discover the information. Providing a template, Providing a step-by-step explanation of the process
Speed, Trust
Check Status S_1_2
Frequent automatic updates
timely response (generated message), that request will be dealt with accordingly (maybe provide the legally allowed response time for the result)
see process for deletion
predictability
Simplicity: E.g. unique link displaying the current status, updates by e-mail or SMS
The software must provide a way to enable this check.
User-friendliness. Clarity.
Status Updates shall be near real time with an audit trail.
Time to log into the system
Trust
Receive Result S_1_3
Receive notification
As in all communication with the data subject, clear and easy understandable language. Ensure response within legally allowed time frame
see process for deletion
consistency
Simplicity: By the same or similar medium as the request for erasure was received (written letter, e-mail, SMS).
The software must provide a way to enable this communication.
User-friendliness. Clarity.
Data Shall by encrypted with the data subject's key. Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.)
How many types of notification are supported?
Speed, Usability
Proof Identity S_2
Easy verification process
clear identification (ID card, passport, driver's licence), if the controller has reasonable doubts about the ID of the data subject. Assurance to data subject that their ID is only used for the identification process and deleted accordingly afterw
see process for deletion
localized
Safety and simplicity: Online via webcam and passport/ID; digital identification (e.g. Buergerkarte), in case of simple online services (non-governmental/-financial) via username and password.
The subject must be willing to share this proof of identity. The software must provide a way proof the subject's identity.
User-friendliness.
Software shall check if proof identity quality is good enough for further processing
Number of prerequisites the user needs to proof this.
Trust
Check Delay Explanation S_6
Provide explanation. Options moving forward.
Clear and full and timely explanation of the reasons of the delay. (Minor remark: in this diagram it's rather about the access to data and not the deletion, right?)
see process for deletion
comprehensible
Simplicity: Overview of process and outcome, reasoning for the non-deletion; possible remedies.
User-friendliness. Understandability.
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.). Also, possibility to get in touch with the processor.
Time to log into the system
Speed
Wait for Delayed Result S_7
Notification of expiry of deadline. Provide (link) form for complaint.
Ensure to be within the legally required response time
see process for deletion
predictability
Simplicity and safety: Information after 8 weeks; auto-generation of a complaint (at this point the id of the person, the issue and the relevant authority should be known).
Easiness.
Complaint shall be automatically filed if "okeyed" initially. I.e. "do you want to automatically file a complaint after 8 weeks of non-resolution? yes/no"
(Shouldn't that happen automatically, e.g. require this process to be implemented at the provider side)
Speed
Check Positive Result S_8
Provide legal requirements for rejection.
Provide clear and concise reasons for the rejection.
see process for deletion. clearness of explanation (what's the criterion? are there templates or pre-defined reasons? if you don't understand, can you ask?)
comprehensible
Simplicity and safety: Information on reasons, possible remedies, and auto-generation of complaint (see S_7).
Understandability.
Full Report with (process) audit rail signed by the data processor.
Check if necessary, information is included in the result.
Speed, Trust
Check Partial Erasure Explanation S_9
Options to save and export data.
Provide not only the data files, but also an understandable explanation for data subjects not so familiar with this topic.
see process for deletion
traceability. Compatibility
User-friendliness. Understandability.
Data shall have human readable filenames that reflect the content of each file. data shall be in an industry standard file format.
Does the explanation contain technical terms?
Usability
File Complaint S_4
Instruction on formal requirements.
Possibility to file the complaint within a few clicks (form) Usability
see process for deletion
fault tolerance for user input
Simplicity and safety: Auto-generation of complaint (see S_7).
User-friendliness. Multilingualism.
Should be automated with a full collection of all steps that happened so far (audit trail)
Is there template to guide through the process.
Trust
Receive Complaint Result S_5
Notification. Provide legal background and solitons moving forward.
Informing about further possibilities, if the result of the compliant is not satisfactory (e.g. appeal against decision of DPA)
see process for deletion
accuracy
Simplicity and safety: Overview of process and outcome, reasoning; information on further remedies against the decision.
User-friendliness. Understandability.
Does it contain instructions to what to do next?
Trust, Usability

Figure D.3: Mindmap of Analysis Step 1

D.3.2 Analysis Step 2: Splitting

The answers are split into criteria and reduced if they are similar.

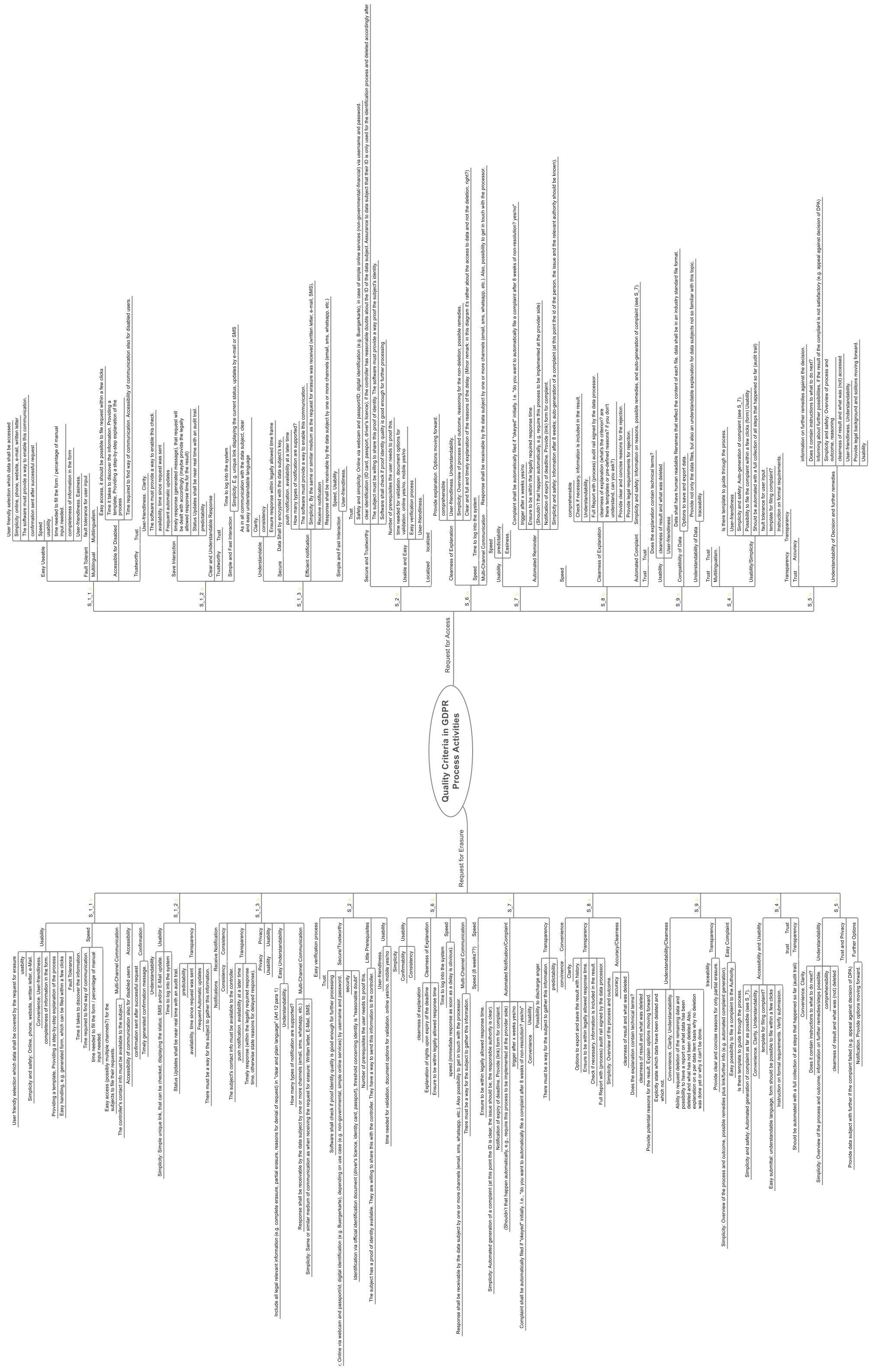

Quality Criteria in GDPR Process Activities
Request for Access
S_1_1
Easy Useable
User friendly selection which data shall be accessed
Simplicity: online, phone, website, e-mail, written letter
The software must provide a way to enable this communication.
confirmation sent after successful request
Speed
usability
time needed to fill the form / percentage of manual input needed.
completeness of information in the form
User-friendliness. Easiness.
Fault Tolerant
fault tolerance for user input
Multilingual
Multilingualism.
Accessible for Disabled
Easy access, it should be possible to file request within a few clicks
Time it takes to discover the information. Providing a template, Providing a step-by-step explanation of the process
Time required to find way of communication. Accessibility of communication also for disabled users.
Trustworthy
Trust
S_1_2
Save Interaction
User-friendliness. Clarity.
The software must provide a way to enable this check.
availability. time since request was sent
Frequent automatic updates
timely response (generated message), that request will be dealt with accordingly (maybe provide the legally allowed response time for the result)
Status Updates shall be near real time with an audit trail.
predictability
Clear and Understandable Response
Trustworthy
Trust
Simple and Fast Interaction
Time to log into the system
Simplicity: E.g. unique link displaying the current status, updates by e-mail or SMS
S_1_3
Understandable
As in all communication with the data subject, clear and easy understandable language
Clarity.
consistency
Ensure response within legally allowed time frame
Secure
Data Shall by encrypted with the data subject's key.
Efficient notification
push notification. availability at a later time
How many types of notification are supported?
The software must provide a way to enable this communication.
Simplicity: By the same or similar medium as the request for erasure was received (written letter, e-mail, SMS).
Receive notification
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.)
Simple and Fast Interaction
Speed, Usability
User-friendliness.
S_2
Secure and Trustworthy
Trust
Safety and simplicity: Online via webcam and passport/ID; digital identification (e.g. Buergerkarte), in case of simple online services (non-governmental/-financial) via username and password.
clear identification (ID card, passport, driver's licence), if the controller has reasonable doubts about the ID of the data subject. Assurance to data subject that their ID is only used for the identification process and deleted accordingly after
The subject must be willing to share this proof of identity. The software must provide a way proof the subject's identity.
Software shall check if proof identity quality is good enough for further processing
Usable and Easy
Number of prerequisites the user needs to proof this.
time needed for validation. document options for validation. online yes/no. mobile yes/no
Easy verification process
User-friendliness.
Localized
localized
S_6
Clearness of Explanation
Provide explanation. Options moving forward.
comprehensible
User-friendliness. Understandability.
Simplicity: Overview of process and outcome, reasoning for the non-deletion; possible remedies.
Clear and full and timely explanation of the reasons of the delay. (Minor remark: in this diagram it's rather about the access to data and not the deletion, right?)
Speed
Time to log into the system
Multi-Channel Communication
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.). Also, possibility to get in touch with the processor.
S_7
Usability
Speed
predictability
Easiness.
Automated Reminder
Complaint shall be automatically filed if "okeyed" initially. I.e. "do you want to automatically file a complaint after 8 weeks of non-resolution? yes/no"
trigger after x weeks yes/no
Ensure to be within the legally required response time
(Shouldn't that happen automatically, e.g. require this process to be implemented at the provider side)
Notification of expiry of deadline. Provide (link) form for complaint.
Simplicity and safety: Information after 8 weeks; auto-generation of a complaint (at this point the id of the person, the issue and the relevant authority should be known).
S_8
Speed
Clearness of Explanation
comprehensible
Check if necessary, information is included in the result.
Understandability.
Full Report with (process) audit rail signed by the data processor.
clearness of explanation (what's the criterion? are there templates or pre-defined reasons? if you don't understand, can you ask?)
Provide clear and concise reasons for the rejection.
Provide legal requirements for rejection.
Automated Complaint
Simplicity and safety: Information on reasons, possible remedies, and auto-generation of complaint (see S_7).
Trust
Trust
S_9
Usability
Does the explanation contain technical terms?
clearness of result and what was deleted
User-friendliness
Compatibility of Data
Data shall have human readable filenames that reflect the content of each file. data shall be in an industry standard file format.
Options to save and export data.
Understandability of Data
Provide not only the data files, but also an understandable explanation for data subjects not so familiar with this topic.
traceability
S_4
Trust
Trust
Multilingualism.
Usability/Simplicity
Is there template to guide through the process.
User-friendliness
Simplicity and safety: Auto-generation of complaint (see S_7).
Possibility to file the complaint within a few clicks (form) Usability
Should be automated with a full collection of all steps that happened so far (audit trail)
fault tolerance for user input
template for filing complaint?
Instruction on formal requirements.
S_5
Transparency
Transparancy
Trust
Accuracy
Understandability of Decision and further remedies
information on further remedies against the decision.
Does it contain instructions to what to do next?
Informing about further possibilities, if the result of the compliant is not satisfactory (e.g. appeal against decision of DPA)
Simplicity and safety: Overview of process and outcome, reasoning
clearness of result and what was (not) accessed
User-friendliness. Understandability.
Provide legal background and solitons moving forward.
Usability
Request for Erasure
S_1_1
Usability
User friendly selection which data shall be covered by the request for erasure
usability
Simplicity and safety: Online, phone, website, written letter, e-Mail.
Convenience. User-friendliness.
completeness of information in the form.
Providing a template, Providing a step-by-step explanation of the process
Easy handling, e.g. generated form, which can be filed within a few clicks
Fault Tolerance
Speed
Time it takes to discover the information.
Time required to find way of communication.
time needed to fill the form / percentage of manual input needed.
Multi-Channel Communication
Easy access (possibly multiple channels?) for the subjects to file their requests.
The controller's contact info must be available to the subject.
Accessibility
Accessibility of communication also for disabled users.
Confirmation
confirmation sent after successful request
Timely generated confirmation message.
S_1_2
Usability
Understandability
Simplicity: Simple unique link, that can be checked, displaying the status; SMS and/or E-Mail update.
Time to log into the system
Status Updates shall be near real time with an audit trail.
Transparency
predictability
availability. time since request was sent
Frequent Automatic updates
There must be a way for the subject to gather this information.
S_1_3
Receive Notification
Notifications
Consistency
Consistency
Transparency
The subject's contact info must be available to the controller.
push notification. availability at a later time
Timely response (within the legally required response time, otherwise state reasons for delayed response).
Privacy
Privacy
Usability
Usability
Easy Understandability
Include all legal relevant information (e.g. complete erasure, partial erasure, reasons for denial of request) in "clear and plain language" (Art 12 para 1)
Understandability.
Multi-Channel Communication
How many types of notification are supported?
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.)
Simplicity: Same or similar medium of communication as when receiving the request for erasure: Written letter, E-Mail, SMS.
S_2
Easy verification process
Secure/Trustworthy
Trust
Software shall check if proof identity quality is good enough for further processing
r; Online via webcam and passport/id; digital identification (e.g. Buergerkarte), depending on use case (e.g. non-governmental, simple online services) by username and password.
security
Identification via official identification document (driver's licence, identity card, passport), threshold concerning identity is "reasonable doubt"
Little Prerequisites
Number of prerequisites the user needs to proof this.
The subject has a proof of identity available. They are willing to share this with the controller. They have a way to send this information to the controller.
Usability
User-friendliness.
time needed for validation. document options for validation. online yes/no. mobile yes/no
S_6
Usability
Simplicity
Confirmability
Consistency
Clearness of Explanation
clearness of explanation
Explanation of rights upon expiry of the deadline
Ensure to be within legally allowed response time
Speed
Time to log into the system
speed (immediate response as soon as a delay is obvious).
Multi-Channel Communication
Response shall be receivable by the data subject by one or more channels (email, sms, whatsapp, etc.). Also possibility to get in touch with the processor.
There must be a way for the subject to gather this information.
S_7
Speed
Speed (8 weeks??)
Automated Notification/Complaint
Ensure to be within legally allowed response time.
Simplicity: Automated generation of a complaint (at this point the ID is clear, the issue should be, the responsible authority is clear).
Notification of expiry of deadline. Provide (link) form for complaint.
(Shouldn't that happen automatically, e.g., require this process to be implemented at the provider side)
trigger after x weeks yes/no
Complaint shall be automatically filed if "okeyed" initially. I.e., "do you want to automatically file a complaint after 8 weeks of non-resolution? yes/no"
Convenience
Usability
Possibility to discharge anger.
Transparency
There must be a way for the subject to gather this information.
predictability
S_8
Convenience
convinience
Transparency
Clarity.
Options to export and save the result with history.
Ensure to be within legally allowed response time.
Check if necessary, information is included in the result.
Full Report with (process) audit rail signed by the data processor.
Simplicity: Overview of the process and outcome.
Accuracy/Clearness
accuracy
clearness of result and what was deleted
S_9
Understandability/Clearness
Does the explanation contain technical terms?
clearness of result and what was deleted
Provide potential reasons for the result. Explain options moving forward.
Explicitly state which data have been deleted and which not.
Convenience. Clarity. Understandability.
Ability to request deletion of the remaining data and possibility to have a report on what data has been deleted and what has not been deleted. Also, explanation on a per data item basis why no deletion was done yet or why it can't be done.
Transparency
traceability
Provide clear and concise reasons for only partial erasure.
Easy Complaint
Simplicity: Overview of the process and outcome, possible remedies plus link/further info (e.g. automated complaint generation).
Easy possibility to file a complaint to the Authority.
S_4
Accessibility and Usability
Is there template to guide through the process.
Simplicity and safety: Automated generation of complaint as far as possible (see S_7).
Convenience. Clarity. Understandability.
template for filing complaint?
Easy submittal: understandable language, form should be possible to file within a few clicks
Instruction on formal requirements. Verify submission.
trust
Trust
Transparency
Should be automated with a full collection of all steps that happened so far (audit trail)
S_5
Understandability
Convenience. Clarity.
Does it contain instructions to what to do next?
Simplicity: Overview of the process and outcome; information on further remedies/steps possible.
comprehensibility
clearness of result and what was (not) deleted
Trust and Privacy
Further Options
Provide data subject with further if the complaint failed (e.g. appeal against decision of DPA)
Notification. Provide options moving forward.

Figure D.4: Mindmap of Analysis Step 2

D.3.3 Analysis Step 3: Question inference

The criteria are sorted and questions are created to ask for the corresponding criteria.

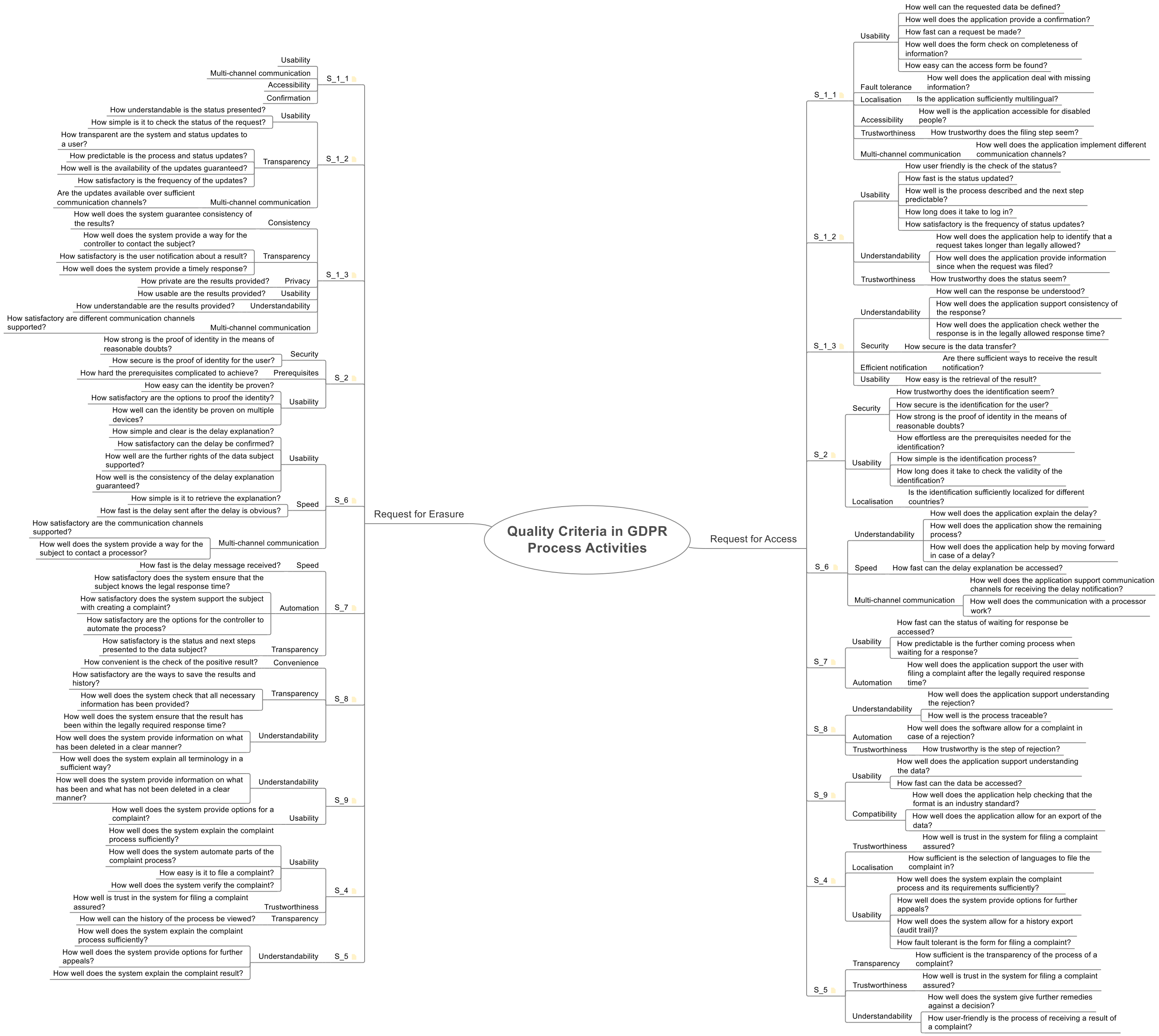

Quality Criteria in GDPR Process Activities
Request for Erasure
S_1_1
Usability
Multi-channel communication
Accessibility
Confirmation
S_1_2
Usability
How understandable is the status presented?
How simple is it to check the status of the request?
Transparency
How transparent are the system and status updates to a user?
How predictable is the process and status updates?
How well is the availability of the updates guaranteed?
How satisfactory is the frequency of the updates?
Multi-channel communication
Are the updates available over sufficient communication channels?
S_1_3
Consistency
How well does the system guarantee consistency of the results?
Transparency
How well does the system provide a way for the controller to contact the subject?
How satisfactory is the user notification about a result?
How well does the system provide a timely response?
Privacy
How private are the results provided?
Usability
How usable are the results provided?
Understandability
How understandable are the results provided?
Multi-channel communication
How satisfactory are different communication channels supported?
S_2
Security
How strong is the proof of identity in the means of reasonable doubts?
How secure is the proof of identity for the user?
Prerequisites
How hard the prerequisites complicated to achieve?
Usability
How easy can the identity be proven?
How satisfactory are the options to proof the identity?
How well can the identity be proven on multiple devices?
S_6
Usability
How simple and clear is the delay explanation?
How satisfactory can the delay be confirmed?
How well are the further rights of the data subject supported?
How well is the consistency of the delay explanation guaranteed?
Speed
How simple is it to retrieve the explanation?
How fast is the delay sent after the delay is obvious?
Multi-channel communication
How satisfactory are the communication channels supported?
How well does the system provide a way for the subject to contact a processor?
S_7
Speed
How fast is the delay message received?
Automation
How satisfactory does the system ensure that the subject knows the legal response time?
How satisfactory does the system support the subject with creating a complaint?
How satisfactory are the options for the controller to automate the process?
Transparency
How satisfactory is the status and next steps presented to the data subject?
S_8
Convenience
How convenient is the check of the positive result?
Transparency
How satisfactory are the ways to save the results and history?
How well does the system check that all necessary information has been provided?
Understandability
How well does the system ensure that the result has been within the legally required response time?
How well does the system provide information on what has been deleted in a clear manner?
S_9
Understandability
How well does the system explain all terminology in a sufficient way?
How well does the system provide information on what has been and what has not been deleted in a clear manner?
Usability
How well does the system provide options for a complaint?
S_4
Usability
How well does the system explain the complaint process sufficiently?
How well does the system automate parts of the complaint process?
How easy is it to file a complaint?
How well does the system verify the complaint?
Trustworthiness
How well is trust in the system for filing a complaint assured?
Transparency
How well can the history of the process be viewed?
S_5
Understandability
How well does the system explain the complaint process sufficiently?
How well does the system provide options for further appeals?
How well does the system explain the complaint result?
Request for Access
S_1_1
Usability
How well can the requested data be defined?
How well does the application provide a confirmation?
How fast can a request be made?
How well does the form check on completeness of information?
How easy can the access form be found?
Fault tolerance
How well does the application deal with missing information?
Localisation
Is the application sufficiently multilingual?
Accessibility
How well is the application accessible for disabled people?
Trustworthiness
How trustworthy does the filing step seem?
Multi-channel communication
How well does the application implement different communication channels?
S_1_2
Usability
How user friendly is the check of the status?
How fast is the status updated?
How well is the process described and the next step predictable?
How long does it take to log in?
How satisfactory is the frequency of status updates?
Understandability
How well does the application help to identify that a request takes longer than legally allowed?
How well does the application provide information since when the request was filed?
Trustworthiness
How trustworthy does the status seem?
S_1_3
Understandability
How well can the response be understood?
How well does the application support consistency of the response?
How well does the application check wether the response is in the legally allowed response time?
Security
How secure is the data transfer?
Efficient notification
Are there sufficient ways to receive the result notification?
Usability
How easy is the retrieval of the result?
S_2
Security
How trustworthy does the identification seem?
How secure is the identification for the user?
How strong is the proof of identity in the means of reasonable doubts?
Usability
How effortless are the prerequisites needed for the identification?
How simple is the identification process?
How long does it take to check the validity of the identification?
Localisation
Is the identification sufficiently localized for different countries?
S_6
Understandability
How well does the application explain the delay?
How well does the application show the remaining process?
How well does the application help by moving forward in case of a delay?
Speed
How fast can the delay explanation be accessed?
Multi-channel communication
How well does the application support communication channels for receiving the delay notification?
How well does the communication with a processor work?
S_7
Usability
How fast can the status of waiting for response be accessed?
How predictable is the further coming process when waiting for a response?
Automation
How well does the application support the user with filing a complaint after the legally required response time?
S_8
Understandability
How well does the application support understanding the rejection?
How well is the process traceable?
Automation
How well does the software allow for a complaint in case of a rejection?
Trustworthiness
How trustworthy is the step of rejection?
S_9
Usability
How well does the application support understanding the data?
How fast can the data be accessed?
Compatibility
How well does the application help checking that the format is an industry standard?
How well does the application allow for an export of the data?
S_4
Trustworthiness
How well is trust in the system for filing a complaint assured?
Localisation
How sufficient is the selection of languages to file the complaint in?
Usability
How well does the system explain the complaint process and its requirements sufficiently?
How well does the system provide options for further appeals?
How well does the system allow for a history export (audit trail)?
How fault tolerant is the form for filing a complaint?
S_5
Transparency
How sufficient is the transparency of the process of a complaint?
Trustworthiness
How well is trust in the system for filing a complaint assured?
Understandability
How well does the system give further remedies against a decision?
How user-friendly is the process of receiving a result of a complaint?

Figure D.5: Mindmap of Analysis Step 3

D.4 Design of Questionnaire Round 2: Assessment of Criteria

Questionnaire round 2 assesses the importance of the identified criteria.

Table D.3: questionnaire 2

Question	**Options**
Q1. I agree to the terms and conditions of this survey.	Accept Decline
Q2. Please sort the following criteria for "Create Request for Erasure" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.	How easy can the deletion form be found? How much time is needed to file the request? How well can the data to be deleted be defined? How well is the process for creating request for erasure explained? How well is the use of templates when creating a request? How well does the form check on completeness of information when creating a request?

Q3. Please sort the following criteria for "Check Status" S_1_2: The data subject checks whether the request has been done.	How simple is it to check the status of the request? How transparent are the system and status updates to a user? How understandable is the status presented? How predictable is the process and status updates? How satisfactory is the frequency of the updates? Are the updates available over sufficient communication channels? How well is the availability of the updates guaranteed?
Q4. Please sort the following criteria for "Receive Result" S_1_3: The data subject receives a response from the data controller.	How understandable are the results provided? How usable are the results provided? How private are the results provided? How satisfactory is the user notification about a result? How well does the system provide a way for the controller to contact the subject? How well does the system provide a timely response? How well does the system guarantee consistency of the results? How satisfactory are different communication channels supported?

Q5. Please sort the following criteria for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.	How secure is the proof of identity for the user? How strong is the proof of identity in the means of reasonable doubts? How hard the prerequisites complicated to achieve? How easy can the identity be proven? How satisfactory are the options to proof the identity? How well can the identity be proven on multiple devices?
Q6. Please sort the following criteria for "Check Delay Explanation" S_6: The data subject receives and checks a notification of a delayed result with an explanation to the data subject.	How simple and clear is the delay explanation? How well are the further rights of the data subject supported? How simple is it to retrieve the explanation? How fast is the delay sent after the delay is obvious? How satisfactory are the communication channels supported? How well does the system provide a way for the subject to contact a processor? How well is the consistency of the delay explanation guaranteed? How satisfactory can the delay be confirmed?

Q7. Please sort the following criteria for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.	How satisfactory does the system ensure that the subject knows the legal response time? How satisfactory does the system support the subject with creating a complaint? How fast is the delay message received? How satisfactory is the status and next steps presented to the data subject? How satisfactory are the options for the controller to automate the process?
Q8. Please sort the following criteria for "Check Positive Result" S_8: The data subject receives and checks a positive result (data erased).	How well does the system ensure that the result has been within the legally required response time? How well does the system provide information on what has been deleted in a clear manner? How well does the system check that all necessary information has been provided? How satisfactory are the ways to save the results and history? How convenient is the check of the positive result?

Q9. Please sort the following criteria for "Check Partial Erasure Explanation" S_9: The data subject receives and checks a result stating that only parts of its data have been deleted.	How well does the system provide options for a complaint? How well does the system explain all terminology in a sufficient way? How well does the system provide information on what has been and what has not been deleted in a clear manner?
Q10. Please sort the following criteria for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.	How easy is it to file a complaint? How well does the system verify the complaint? How well does the system automate parts of the complaint process? How well does the system explain the complaint process sufficiently? How well is trust in the system for filing a complaint assured? How well can the history of the process be viewed?
Q11. Please sort the following criteria for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.	How well does the system explain the complaint process sufficiently? How well does the system explain the complaint result? How well does the system provide options for further appeals?

Q12. Please sort the following criteria for "Create Request for Access" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.	How well can the requested data be defined? How easy can the access form be found? How fast can a request be made? How well does the form check on completeness of information? How well does the application provide a confirmation? How well does the application deal with missing information? How well is the application accessible for disabled people? Is the application sufficiently multilingual? How trustworthy does the filing step seem? How well does the application implement different communication channels?

Q13. Please sort the following criteria for "Check Status" S_1_2: The data subject checks whether the request has been done.	How user friendly is the check of the status? How satisfactory is the frequency of status updates? How well does the application help to identify that a request takes longer than legally allowed? How fast is the status updated? How well is the process described and the next step predictable? How long does it take to log in? How well does the application provide information since when the request was filed? How trustworthy does the status seem?
Q14. Please sort the following criteria for "Receive Result" S_1_3: The data subject receives a response from the data controller.	How well can the response be understood? How secure is the data transfer? How easy is the retrieval of the result? How well does the application check wether the response is in the legally allowed response time? How well does the application support consistency of the response? Are there sufficient ways to receive the result notification?

Q15. Please sort the following criteria for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.	How simple is the identification process? How secure is the identification for the user? How strong is the proof of identity in the means of reasonable doubts? How effortless are the prerequisites needed for the identification? How trustworthy does the identification seem? How long does it take to check the validity of the identification? Is the identification sufficiently localized for different countries?
Q16. Please sort the following criteria for "Check Delay Explanation" S_6: The data subject receives and checks a result stating that only parts of its data have been accessed.	How well does the application explain the delay? How fast can the delay explanation be accessed? How well does the application show the remaining process? How well does the application help by moving forward in case of a delay? How well does the application support communication channels for receiving the delay notification? How well does the communication with a processor work?

Q17. Please sort the following criteria for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.	How fast can the status of waiting for response be accessed? How predictable is the further coming process when waiting for a response? How well does the application support the user with filing a complaint after the legally required response time?
Q18. Please sort the following criteria for "Check Rejection" S_8: If the response was a rejection, the data subject checks the reason for the rejection.	How well does the software allow for a complaint in case of a rejection? How well does the application support understanding the rejection? How trustworthy is the step of rejection? How well is the process traceable?
Q19. Please sort the following criteria for "Open and Understand Received Data" S_9: After receiving the data files, the data subject tries to open and understand the data.	How fast can the data be accessed? How well does the application allow for an export of the data? How well does the application help checking that the format is an industry standard? How well does the application support understanding the data?

Q20. Please sort the following criteria for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.	How well does the system explain the complaint process and its requirements sufficiently? How fault tolerant is the form for filing a complaint? How well does the system allow for a history export (audit trail)? How well is trust in the system for filing a complaint assured? How well does the system provide options for further appeals? How sufficient is the selection of languages to file the complaint in?
Q21. Please sort the following criteria for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.	How user-friendly is the process of receiving a result of a complaint? How well does the system give further remedies against a decision? How sufficient is the transparency of the process of a complaint? How well is trust in the system for filing a complaint assured?

D.5 Results of Questionnaire Round 2: Assessment of Criteria

The raw results of the second round of the questionnaire cannot be depicted properly and are therefore only available in the TU archive. The aggregated results can be found in the analysis chapter.

D.6 Analysis of Questionnaire Round 2: Assessment of Criteria

Questionnaire round 2, question 1

Q1. I agree to the terms and conditions of this survey.	
Options	**Result**
Accept	100%
Decline	0%

Questionnaire round 2, question 2

Q2. Please sort the following criteria for "Create Request for Erasure" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.	
Options	**Score**
How easy can the deletion form be found?	5,7
How much time is needed to file the request?	5,1
How well can the data to be deleted be defined?	3,7
How well is the process for creating request for erasure explained?	3,2
How well is the use of templates when creating a request?	1,8
How well does the form check on completeness of information when creating a request?	1,5

Questionnaire round 2, question 3

Q3. Please sort the following criteria for "Check Status" S_1_2: The data subject checks whether the request has been done.	
Options	**Score**
How simple is it to check the status of the request?	6,9
How transparent are the system and status updates to a user?	6
How understandable is the status presented?	5,1
How predictable is the process and status updates?	3,9
How satisfactory is the frequency of the updates?	3,1

Are the updates available over sufficient communication channels?	1,9
How well is the availability of the updates guaranteed?	1,1

Questionnaire round 2, question 4

Q4. Please sort the following criteria for "Receive Result" S_1_3: The data subject receives a response from the data controller.	
Options	**Score**
How understandable are the results provided?	7,8
How usable are the results provided?	7
How private are the results provided?	5,8
How satisfactory is the user notification about a result?	5,3
How well does the system provide a way for the controller to contact the subject?	4,1
How well does the system provide a timely response?	2,9
How well does the system guarantee consistency of the results?	1,9
How satisfactory are different communication channels supported?	1,2

Questionnaire round 2, question 5

Q5. Please sort the following criteria for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.	
Options	**Score**
How secure is the proof of identity for the user?	5,5
How strong is the proof of identity in the means of reasonable doubts?	4,7
How hard the prerequisites complicated to achieve?	3,5
How easy can the identity be proven?	3,5
How satisfactory are the options to proof the identity?	2,2
How well can the identity be proven on multiple devices?	1,6

Questionnaire round 2, question 6

Q6. Please sort the following criteria for "Check Delay Explanation" S_6: The data subject receives and checks a notification of a delayed result with an explanation to the data subject.	
Options	**Score**
How simple and clear is the delay explanation?	7,9
How well are the further rights of the data subject supported?	6,6
How simple is it to retrieve the explanation?	5,9
How fast is the delay sent after the delay is obvious?	4,9

How satisfactory are the communication channels supported?	4,4
How well does the system provide a way for the subject to contact a processor?	2,9
How well is the consistency of the delay explanation guaranteed?	2,4
How satisfactory can the delay be confirmed?	1

Questionnaire round 2, question 7

Q7. Please sort the following criteria for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.	
Options	**Score**
How satisfactory does the system ensure that the subject knows the legal response time?	4,9
How satisfactory does the system support the subject with creating a complaint?	4
How fast is the delay message received?	3,1
How satisfactory is the status and next steps presented to the data subject?	1,9
How satisfactory are the options for the controller to automate the process?	1,1

Questionnaire round 2, question 8

Q8. Please sort the following criteria for "Check Positive Result" S_8: The data subject receives and checks a positive result (data erased).	
Options	**Score**
How well does the system ensure that the result has been within the legally required response time?	4,6
How well does the system provide information on what has been deleted in a clear manner?	4
How well does the system check that all necessary information has been provided?	2,8
How satisfactory are the ways to save the results and history?	1,8
How convenient is the check of the positive result?	1,8

Questionnaire round 2, question 9

Q9. Please sort the following criteria for "Check Partial Erasure Explanation" S_9: The data subject receives and checks a result stating that only parts of its data have been deleted.	
Options	**Score**

How well does the system provide options for a complaint?	2,9
How well does the system explain all terminology in a sufficient way?	1,8
How well does the system provide information on what has been and what has not been deleted in a clear manner?	1,3

Questionnaire round 2, question 10

Q10. Please sort the following criteria for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.	
Options	**Score**
How easy is it to file a complaint?	6
How well does the system verify the complaint?	4,8
How well does the system automate parts of the complaint process?	3,7
How well does the system explain the complaint process sufficiently?	2,8
How well is trust in the system for filing a complaint assured?	1,9
How well can the history of the process be viewed?	1,8

Questionnaire round 2, question 11

Q11. Please sort the following criteria for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.	
Options	**Score**
How well does the system explain the complaint process sufficiently?	2,9
How well does the system explain the complaint result?	2
How well does the system provide options for further appeals?	1,1

Questionnaire round 2, question 12

Q12. Please sort the following criteria for "Create Request for Access" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.	
Options	**Score**
How well can the requested data be defined?	9,4
How easy can the access form be found?	9
How fast can a request be made?	8,4
How well does the form check on completeness of information?	6,7
How well does the application provide a confirmation?	5,5
How well does the application deal with missing information?	4,8
How well is the application accessible for disabled people?	3,8
How trustworthy does the filing step seem?	3,2

Is the application sufficiently multilingual?	2,6
How well does the application implement different communication channels?	1,6

Questionnaire round 2, question 13

Q13. Please sort the following criteria for "Check Status" S_1_2: The data subject checks whether the request has been done.	
Options	**Score**
How user friendly is the check of the status?	7,8
How satisfactory is the frequency of status updates?	6,7
How well does the application help to identify that a request takes longer than legally allowed?	5,7
How fast is the status updated?	5,3
How well is the process described and the next step predictable?	3,8
How long does it take to log in?	3,4
How well does the application provide information since when the request was filed?	1,8
How trustworthy does the status seem?	1,5

Questionnaire round 2, question 14

Q14. Please sort the following criteria for "Receive Result" S_1_3: The data subject receives a response from the data controller.	
Options	**Score**
How well can the response be understood?	5,7
How secure is the data transfer?	4,8
How easy is the retrieval of the result?	4,5
How well does the application check wether the response is in the legally allowed response time?	2,8
How well does the application support consistency of the response?	2
Are there sufficient ways to receive the result notification?	1,2

Questionnaire round 2, question 15

Q15. Please sort the following criteria for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.	
Options	**Score**
How simple is the identification process?	6,9
How secure is the identification for the user?	5,9
How strong is the proof of identity in the means of reasonable doubts?	4,9

How effortless are the prerequisites needed for the identification?	4,1
How trustworthy does the identification seem?	3
How long does it take to check the validity of the identification?	2,2
Is the identification sufficiently localized for different countries?	1

Questionnaire round 2, question 16

Q16. Please sort the following criteria for "Check Delay Explanation" S_6: The data subject receives and checks a result stating that only parts of its data have been accessed.	
Options	**Score**
How well does the application explain the delay?	6
How fast can the delay explanation be accessed?	5
How well does the application show the remaining process?	4
How well does the application help by moving forward in case of a delay?	2,9
How well does the application support communication channels for receiving the delay notification?	2,1
How well does the communication with a processor work?	1

Questionnaire round 2, question 18

Q18. Please sort the following criteria for "Check Rejection" S_8: If the response was a rejection, the data subject checks the reason for the rejection.	
Options	**Score**
How well does the software allow for a complaint in case of a rejection?	3,8
How well does the application support understanding the rejection?	3,2
How trustworthy is the step of rejection?	1,9
How well is the process traceable?	1,1

Questionnaire round 2, question 19

Q19. Please sort the following criteria for "Open and Understand Received Data" S_9: After receiving the data files, the data subject tries to open and understand the data.	
Options	**Score**
How fast can the data be accessed?	3,8
How well does the application allow for an export of the data?	3
How well does the application help checking that the format is an industry standard?	1,9
How well does the application support understanding the data?	1,3

Questionnaire round 2, question 20

Q20. Please sort the following criteria for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.	
Options	**Score**
How well does the system explain the complaint process and its requirements sufficiently?	6
How fault tolerant is the form for filing a complaint?	4,8
How well does the system allow for a history export (audit trail)?	4,1
How well is trust in the system for filing a complaint assured?	3
How well does the system provide options for further appeals?	2,1
How sufficient is the selection of languages to file the complaint in?	1

Questionnaire round 2, question 21

Q21. Please sort the following criteria for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.	
Options	**Score**
How user-friendly is the process of receiving a result of a complaint?	3,9
How well does the system give further remedies against a decision?	2,9
How sufficient is the transparency of the process of a complaint?	2
How well is trust in the system for filing a complaint assured?	1,2

APPENDIX E

Appendix: Expert Interviews Regarding Data Protection Implementations

The following sections show the design, results and detailed analysis of the expert interviews done in order to grade the applications.

E.1 Design of Questionnaire of the Expert Interview

Question 1 The following questionnaire contains an introduction to two GDPR Processes and questions regarding these processes. The aim is to find criteria to evaluate software trying to help the particular process steps.

We want to inform you about the processing of the data: - The data are subjected to statistical analyses. - They are used and published in aggregated form for publications. - The data will be passed on to other researchers for further research. - The data will be archived in a secure repository for the long term. I have been informed in writing about the procedure for the collection, data storage and analysis of the answers I have given. I am aware that participation in this interview is voluntary and that I have the option to cancel the interview at any time.

Options:

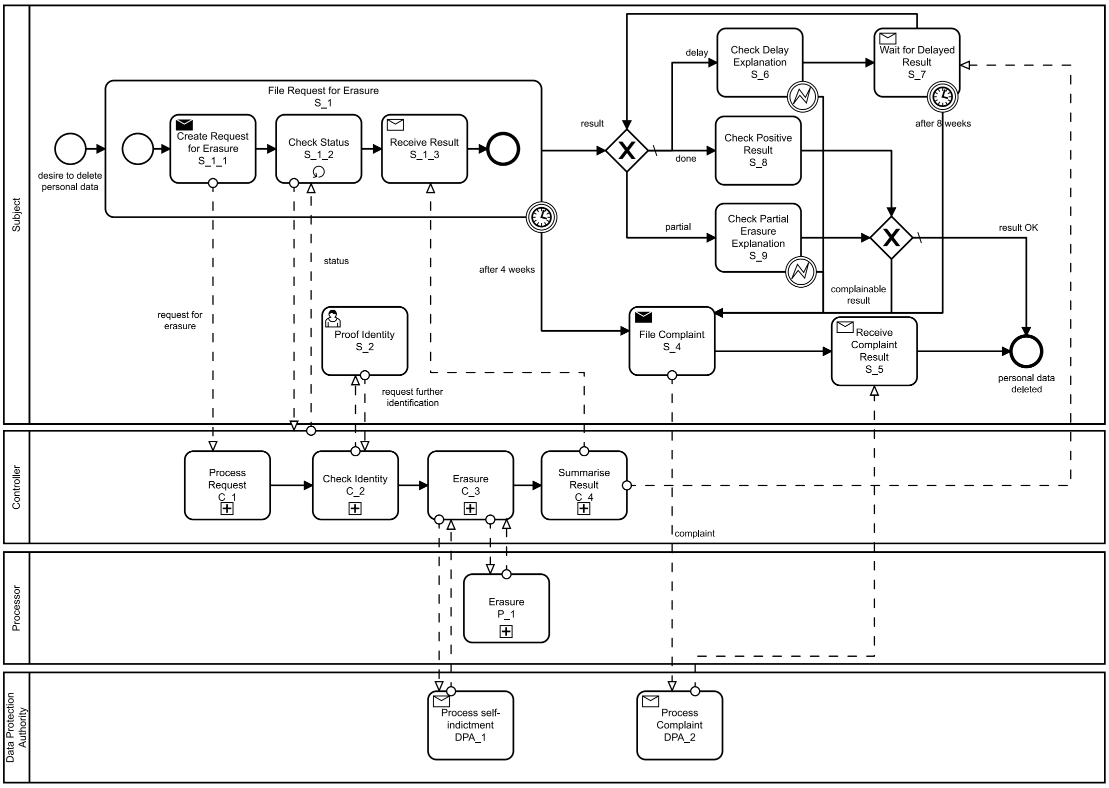
- Accept
- Decline

E.1.1 Explanation of the Process: Request for Erasure

A request to be 'forgotten' or 'request for erasure' is a request a natural person (called data subject) can file against a party that is processing its personal data and decides the purposes and means of the processing (called a controller) of the natural person according to the GDPR (Art. 17). The controller and processor then need to delete the data (with some exceptions).

The following diagram shows the process of filing such a request against a controller that has one subcontractor (processor) who also processes the data on behalf of the controller.

Optional annotation information The horizontal 'lanes' show the person or role executing the activity (each block). The white circle is the start and the bold white circle the end. The arrows show the 'flow' of execution whereas the 'x' symbol in the diamond shows an 'exclusive' gateway, meaning that only one of the following activities are executed. The clock symbol shows a waiting time and the 'lightning' an error. The filled envelop marks an activity that sends a message – the white envelop receives a message.



Next, we want to evaluate an application with the previously found criteria. The first of two applications is an application to file a request for erasure.

E.1.2 Description of the Application to file a Request for Erasure

The application for filing a request for erasure with a controller is a website available on the secure world wide web. To access the website, no login or registration is needed. The request starts with an explanation of the user's rights (the user is called 'data subject' in this context) and the process of filing an erasure request.

The request is created with a wizard, meaning the user is guided through the creation process. The steps of the wizard are depicted on the top of the webpage. Within the request form question mark symbols show further information regarding it.

Initially, the data subject can choose a company or enter their contact information manually. The fields of the form match the corresponding PDF form of the Austrian Data Protection Authority (DPA).

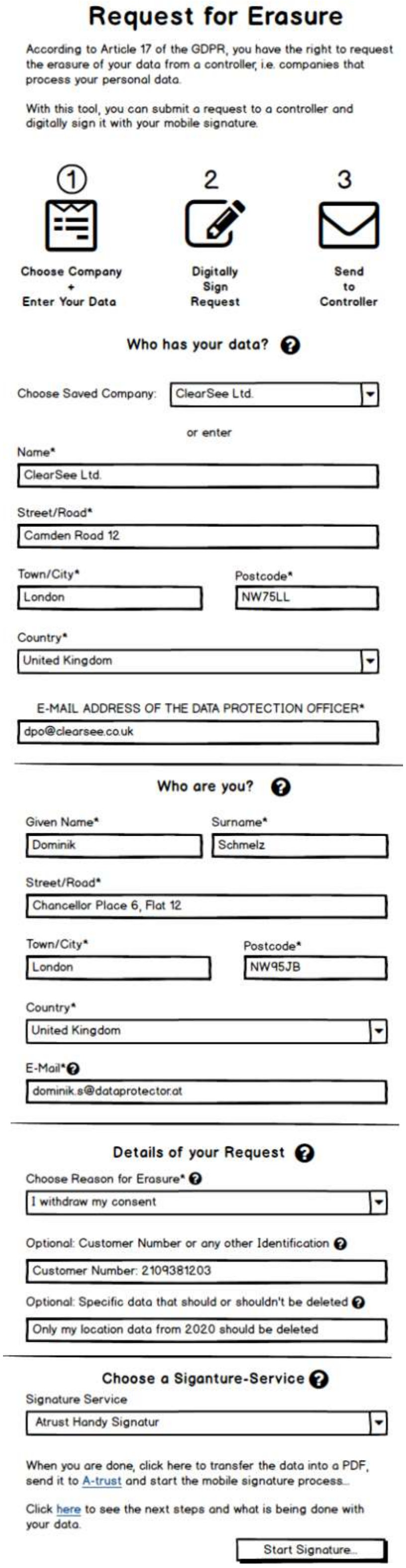
Technical detail: The contact information is manually managed by an administrator.

The data subject then enters its contact information and needs to specify a reason for the erasure request. The requirement for this information has been derived from the required fields in the request form of the Austrian data protection authority.

The data subject can specify an identity known to the controller, such as a user identification or customer number or any other means of recognition for the controller, to allow the controller to select the data of the data subject within its database.

Before submitting the form, the data subject receives an explanation of what happens with the provided data and the next steps.

Technical detail: The information provided by the data subject is put in the PDF mentioned above form of the Austrian data protection authority and transmitted to the A-trust servers as a signature request. The A-trust server returns a URL that must be given to the data subject to authorize the qualified signature.



Request for Erasure

According to Article 17 of the GDPR, you have the right to request the erasure of your data from a controller, i.e. companies that process your personal data.

With this tool, you can submit a request to a controller and digitally sign it with your mobile signature.

Who has your data?

Choose Saved Company: ClearSee Ltd.

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER*: dpo@clearsee.co.uk

Who are you?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom

E-Mail*: dominik.s@dataprotector.at

Details of your Request

Choose Reason for Erasure*: I withdraw my consent

Optional: Customer Number or any other Identification: Customer Number: 2109381203

Optional: Specific data that should or shouldn't be deleted: Only my location data from 2020 should be deleted

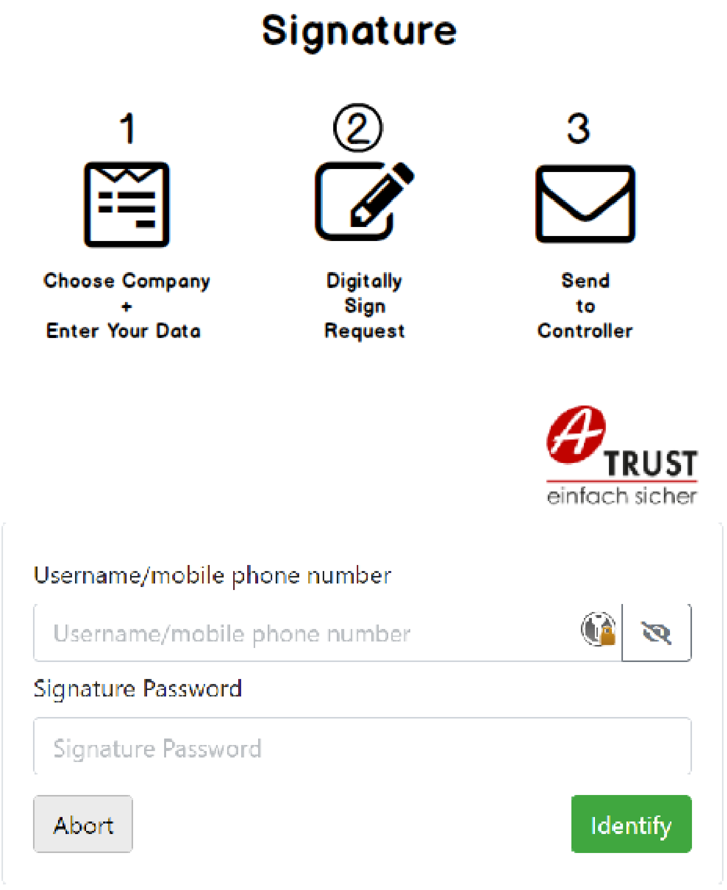
Choose a Siganture-Service

Signature Service: Atrust Handy Signatur

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Click here to see the next steps and what is being done with your data.

After submitting the information, the data subject must authenticate with the chosen (eIDAS) authentication service. In this case its Handy-Signatur (National mobile identification of Austria). After the two factor authentication, the data subject authorizes the qualified signature of the PDF.



Signature

1 Choose Company + Enter Your Data

② Digitally Sign Request

3 Send to Controller

Username/mobile phone number

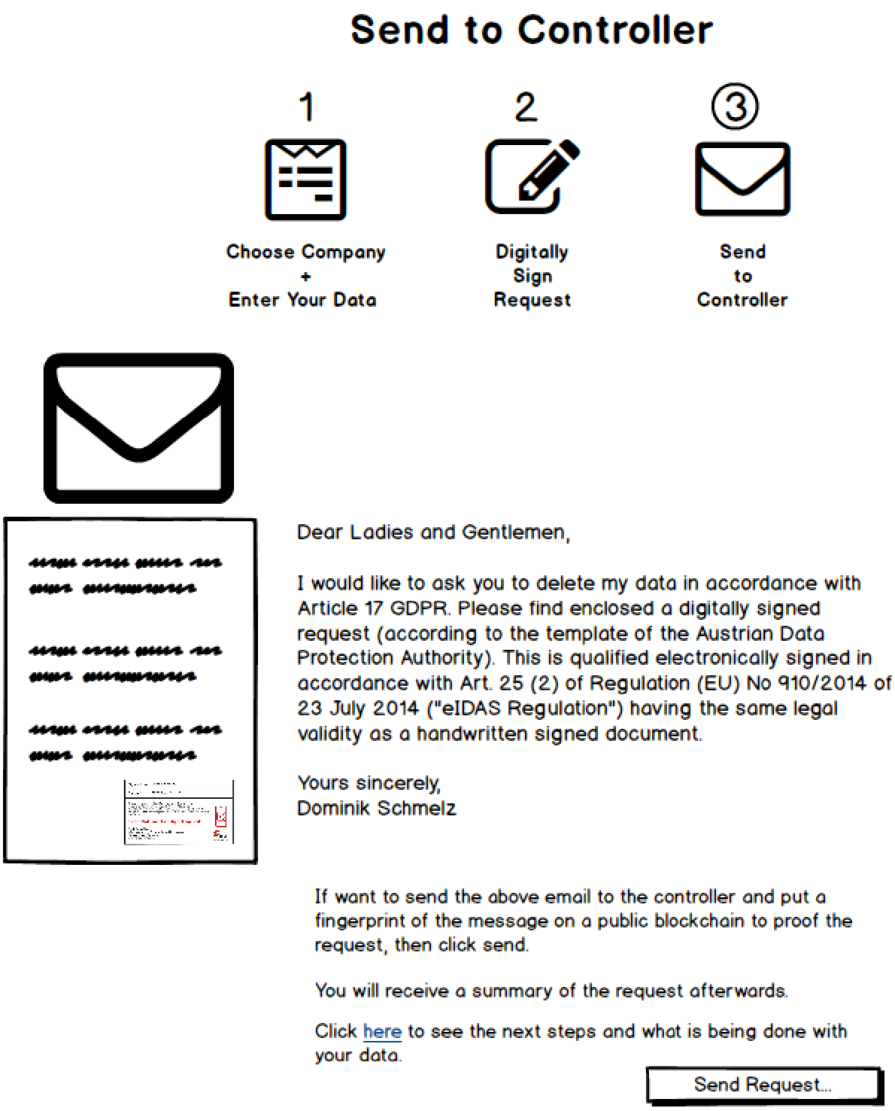
Username/mobile phone number

Signature Password

Signature Password

A summary is shown with the signed PDF, and the email sent to the controller. If the data subject agrees to send the email and the PDF attached, the mail is sent to the controller, and a fingerprint is be stored on a public blockchain.

Technical Detail: The email is sent to the system's own SMTP Server, and the request (including the mail and PDF) is hashed and put on a public blockchain. The free OpenTimestamps service is used for this service. It timestamps the hash value, puts it in a Merkle hash tree, and stores the result on the Bitcoin blockchain. Therefore, no personal data is stored on the blockchain, and the existence of the request can be proven later.



Dear Ladies and Gentlemen,

I would like to ask you to delete my data in accordance with Article 17 GDPR. Please find enclosed a digitally signed request (according to the template of the Austrian Data Protection Authority). This is qualified electronically signed in accordance with Art. 25 (2) of Regulation (EU) No 910/2014 of 23 July 2014 ("eIDAS Regulation") having the same legal validity as a handwritten signed document.

Yours sincerely,
Dominik Schmelz

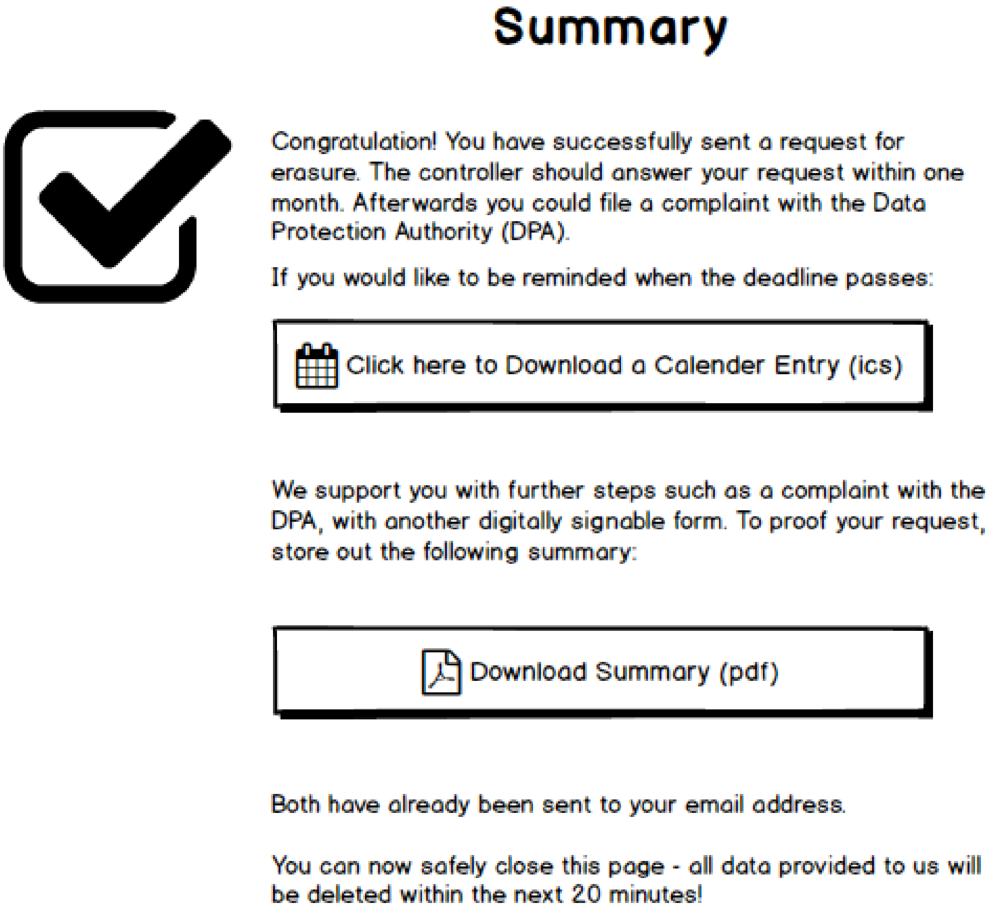
If want to send the above email to the controller and put a fingerprint of the message on a public blockchain to proof the request, then click send.

You will receive a summary of the request afterwards.

Click here to see the next steps and what is being done with your data.

The data subject receives a success message on the webpage and confirmation via email. The message and email include a summary of the request as a PDF and a calendar entry (ICS file) to be stored in the data subjects calendar software. The calendar entry reminds the data subject of the date it can file a complaint with the data protection authority if the controller has not answered the request. In case of a delay, the data subject must receive a notification from the controller; in this case, the data subject must move the calendar entry manually.

The calendar entry includes a link to a form to fill out the complaint information and send it again digitally signed to the Data Protection Authority.



Summary

Congratulation! You have successfully sent a request for erasure. The controller should answer your request within one month. Afterwards you could file a complaint with the Data Protection Authority (DPA).

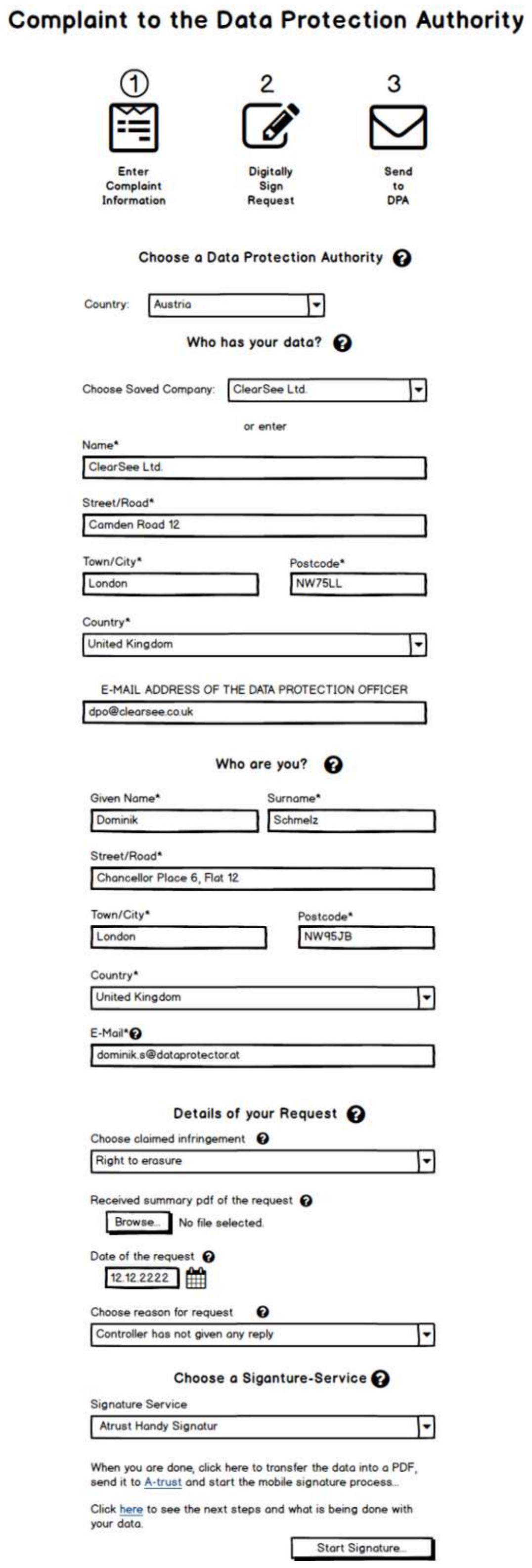
If you would like to be reminded when the deadline passes:

We support you with further steps such as a complaint with the DPA, with another digitally signable form. To proof your request, store out the following summary:

Both have already been sent to your email address.

You can now safely close this page - all data provided to us will be deleted within the next 20 minutes!

To file a complaint, the data subject can click the link provided in the calendar entry. The data subject can choose a data protection authority it wants to address. It must specify the details of the request. These have been derived from the Austrian data protection Authorities complaint form. The summary received via email can be attached to the complaint to prove the request to the authority. Many fields of the complaint form are automatically filled out (such as type of signature, type of contact etc.).



Complaint to the Data Protection Authority

Choose a Data Protection Authority ?

Country: Austria ▾

Who has your data? ?

Choose Saved Company: ClearSee Ltd. ▾

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom ▾

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER: dpo@clearsee.co.uk

Who are you? ?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom ▾

E-Mail* ?: dominik.s@dataprotector.at

Details of your Request ?

Choose claimed infringement ?: Right to erasure ▾

Received summary pdf of the request ?: Browse... No file selected.

Date of the request ?: 12.12.2222

Choose reason for request ?: Controller has not given any reply ▾

Choose a Siganture-Service ?

Signature Service: Atrust Handy Signatur ▾

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Click here to see the next steps and what is being done with your data.

Start Signature...

E.1.3 Evaluation of the Application to file a Request for Erasure

Now please evaluate the application by answering the following questions. If the proposed solution does not support the step (question) at all or the description does not mention a support for the step select "not supportive".

Question 2: Please answer the following criteria questions for "Create Request for Erasure" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.

- How easy can the deletion form be found?
- How much time is needed to file the request?
- How well can the data to be deleted be defined?
- How well is the process for creating request for erasure explained?
- How well is the use of templates when creating a request?
- How well does the form check on completeness of information when creating a request?

Question 3: Please answer the following criteria questions for "Check Status" S_1_2: The data subject checks whether the request has been done.

- How simple is it to check the status of the request?
- How transparent are the system and status updates to a user?
- How understandable is the status presented?
- How predictable is the process and status updates?
- How satisfactory is the frequency of the updates?
- Are the updates available over sufficient communication channels?
- How well is the availability of the updates guaranteed?

Question 4: Please answer the following criteria questions for "Receive Result" S_1_3: The data subject receives a response from the data controller.

- How understandable are the results provided?
- How usable are the results provided?
- How private are the results provided?
- How satisfactory is the user notification about a result?
- How well does the system provide a way for the controller to contact the subject?
- How well does the system provide a timely response?
- How well does the system guarantee consistency of the results?
- How satisfactory are different communication channels supported?

Question 5: Please answer the following criteria questions for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.

- How secure is the proof of identity for the user?

- How strong is the proof of identity in the means of reasonable doubts?
- How hard are are the prerequisites to achieve?
- How easy can the identity be proven?
- How satisfactory are the options to proof the identity?
- How well can the identity be proven on multiple devices?

Question 6: Please answer the following criteria questions for "Check Delay Explanation" S_6: The data subject receives and checks a notification of a delayed result with an explanation to the data subject.

- How simple and clear is the delay explanation?
- How well are the further rights of the data subject supported?
- How simple is it to retrieve the explanation?
- How fast is the delay sent after the delay is obvious?
- How satisfactory are the communication channels supported?
- How well does the system provide a way for the subject to contact a processor?
- How well is the consistency of the delay explanation guaranteed?
- How satisfactory can the delay be confirmed?

Question 7: Please answer the following criteria questions for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.

- How satisfactory does the system ensure that the subject knows the legal response time?
- How satisfactory does the system support the subject with creating a complaint?
- How fast is the delay message received?
- How satisfactory is the status and next steps presented to the data subject?
- How satisfactory are the options for the controller to automate the process?

Question 8: Please answer the following criteria questions for "Check Positive Result" S_8: The data subject receives and checks a positive result (data erased).

- How well does the system ensure that the result has been within the legally required response time?
- How well does the system provide information on what has been deleted in a clear manner?
- How well does the system check that all necessary information has been provided?
- How satisfactory are the ways to save the results and history?
- How convenient is the check of the positive result?

Question 9: Please answer the following criteria questions for "Check Partial Erasure Explanation" S_9: The data subject receives and checks a result stating that only parts of its data have been deleted.

- How well does the system provide options for a complaint?
- How well does the system explain all terminology in a sufficient way?
- How well does the system provide information on what has been and what has not been deleted in a clear manner?

Question 10: Please answer the following criteria questions for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.

- How easy is it to file a complaint?
- How well does the system verify the complaint?
- How well does the system automate parts of the complaint process?
- How well does the system explain the complaint process sufficiently?
- How well is trust in the system for filing a complaint assured?
- How well can the history of the process be viewed?

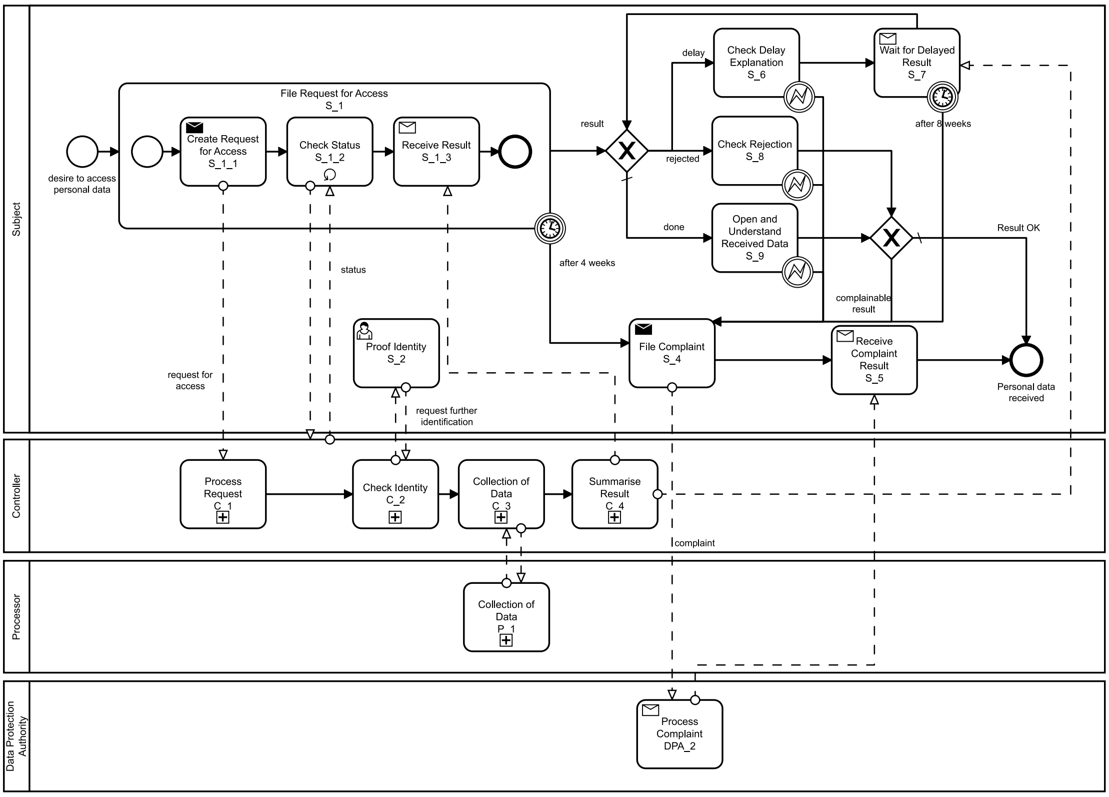
Question 11: Please answer the following criteria questions for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.

- How well does the system explain the complaint process sufficiently?
- How well does the system explain the complaint result?
- How well does the system provide options for further appeals?

E.1.4 Explanation of the Process: Request for Access

A ‘request to access’ is a request a data subject can file against a controller. The controller must then deliver information regarding the processing activities, the rights of the data subject and the data itself.

The following diagram shows the process of filing such a request against a controller that has one processor.



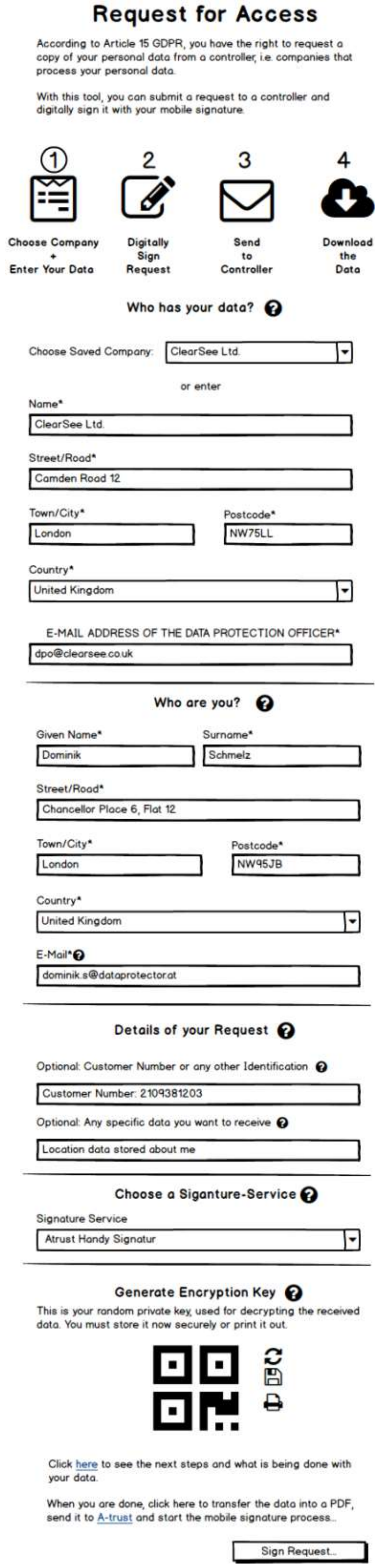
Next, we want to evaluate an application with the previously found criteria. The first of two applications is an application to file a request for erasure.

E.1.5 Description of the Application to file a Request for Access

The application for filing a request for access with a controller is a website available on the open Internet (https). To access the website, no login or registration is needed. The request starts with an explanation of the user's rights (the user is called "data subject" in this context) and the process of filing a request for access.

The request is created with a wizard, meaning the user is guided through the creation process. The steps of the wizard are depicted on the top. The request form shows further information regarding the topic as help icons (question marks). Initially, the data subject can choose a company or enter their contact information manually. The fields of the form are derived from the corresponding form of the Austrian Data Protection Authority (DPA). Technical detail: The contact information is manually managed by an administrator.

The data subject can specify an identity known to the controller, such as a user identification or customer number or any other means of recognition for the controller to allow it to select the data of the data subject within its database. The data subjects web browser automatically generates an encryption key. This key is used to transfer the data securely. It guarantees that only the data subject can read the data. It is never sent to the application's server. Therefore, the service provider of the application cannot access the data. Technical Details: The browser generates a private-public key pair with javascript. Only the public key and a fingerprint (hash) are sent to the server allowing the data subject to only send the fingerprint to the controller (see later in the process). This key pair is only used for one transaction. Before submitting the form, the data subject receives an explanation of what happens with the provided data and the next steps.



Request for Access

According to Article 15 GDPR, you have the right to request a copy of your personal data from a controller, i.e. companies that process your personal data.

With this tool, you can submit a request to a controller and digitally sign it with your mobile signature.

Who has your data?

Choose Saved Company: ClearSee Ltd.

or enter

Name*
ClearSee Ltd.

Street/Road*
Camden Road 12

Town/City*
London

Postcode*
NW75LL

Country*
United Kingdom

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER*
dpo@clearsee.co.uk

Who are you?

Given Name*
Dominik

Surname*
Schmelz

Street/Road*
Chancellor Place 6, Flat 12

Town/City*
London

Postcode*
NW95JB

Country*
United Kingdom

E-Mail*
dominik.s@dataprotector.at

Details of your Request

Optional: Customer Number or any other Identification
Customer Number: 2109381203

Optional: Any specific data you want to receive
Location data stored about me

Choose a Siganture-Service

Signature Service
Atrust Handy Signatur

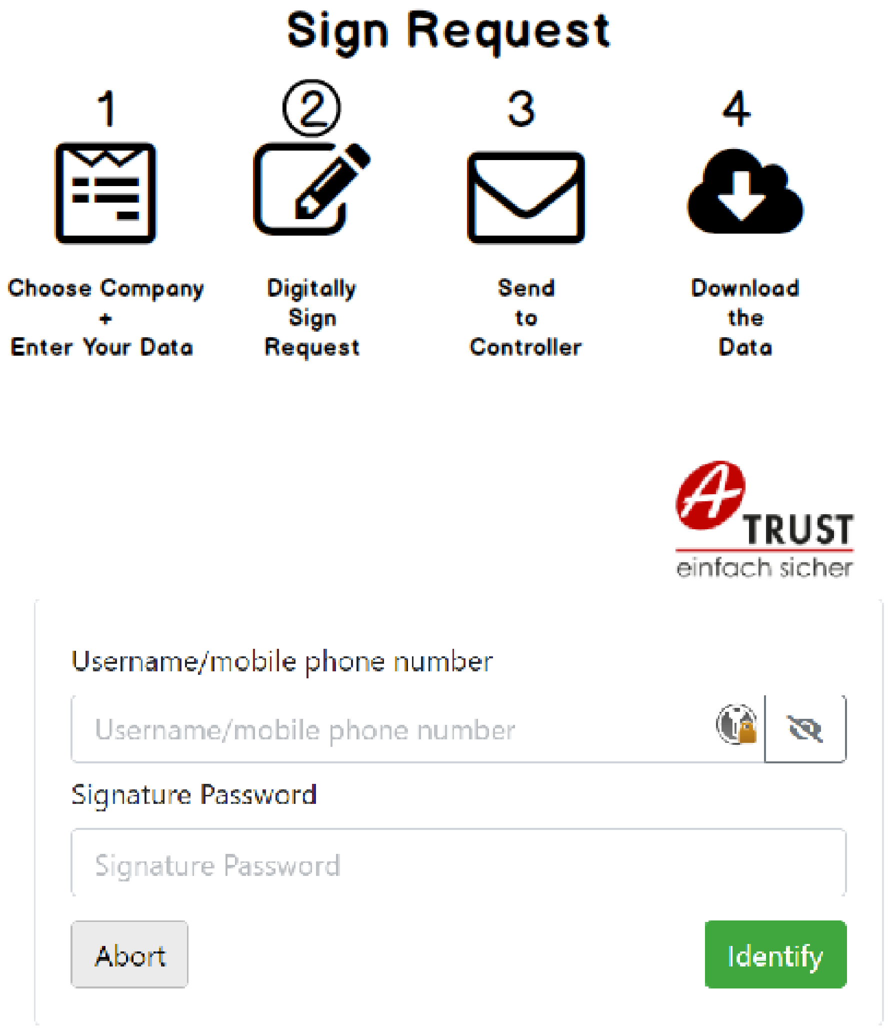
Generate Encryption Key

This is your random private key, used for decrypting the received data. You must store it now securely or print it out.

Click here to see the next steps and what is being done with your data.

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

After submitting the information, the data subject must authenticate with the chosen (eIDAS) authentication service. In this case its Handy-Signatur (National mobile identification of Austria). After the two factor authentication, the data subject authorizes the qualified signature of the PDF.



Sign Request

1 Choose Company + Enter Your Data

② Digitally Sign Request

3 Send to Controller

4 Download the Data

Username/mobile phone number

Username/mobile phone number

Signature Password

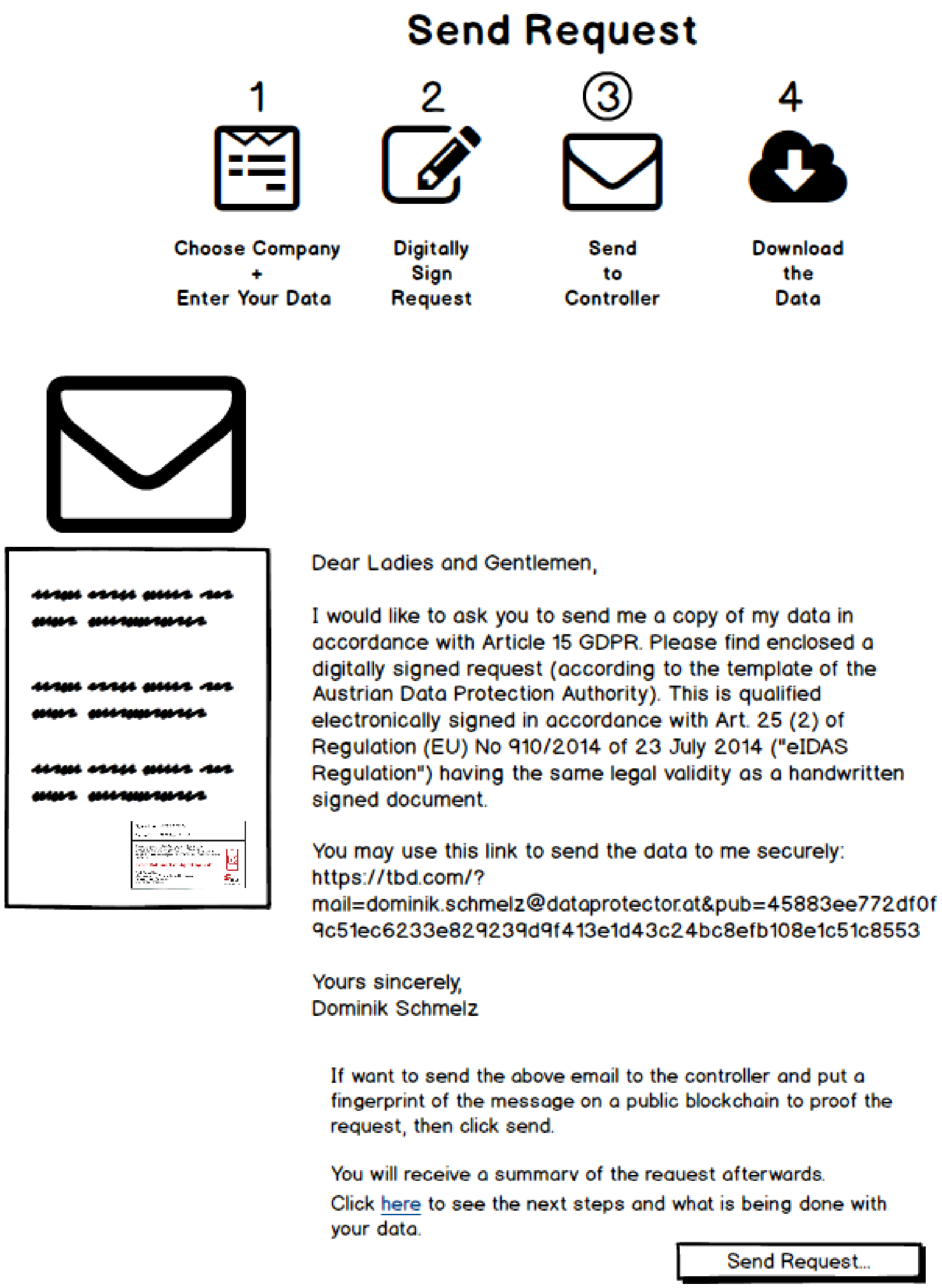
Signature Password

A summary is shown with the signed PDF, and the email is sent to the controller. If the data subject agrees to send the email and the PDF attached, the mail will be sent to the controller, and a fingerprint will be stored on a public blockchain.

Technical Detail: The email is sent to the system's own SMTP Server, and the request (including the mail and PDF) is hashed and put on a public blockchain. The free OpenTimestamps service is used for this service. It timestamps the hash value, puts it in a Merkle hash tree, and stores the result on the Bitcoin blockchain. Therefore, no personal data is stored on the blockchain, and the existence of the request can be proven later.

The mail to the controller contains a link it can use to upload the data for the data subject.

Technical Detail: The link contains the email address and the fingerprint of the public key of the data subject.



Dear Ladies and Gentlemen,

I would like to ask you to send me a copy of my data in accordance with Article 15 GDPR. Please find enclosed a digitally signed request (according to the template of the Austrian Data Protection Authority). This is qualified electronically signed in accordance with Art. 25 (2) of Regulation (EU) No 910/2014 of 23 July 2014 ("eIDAS Regulation") having the same legal validity as a handwritten signed document.

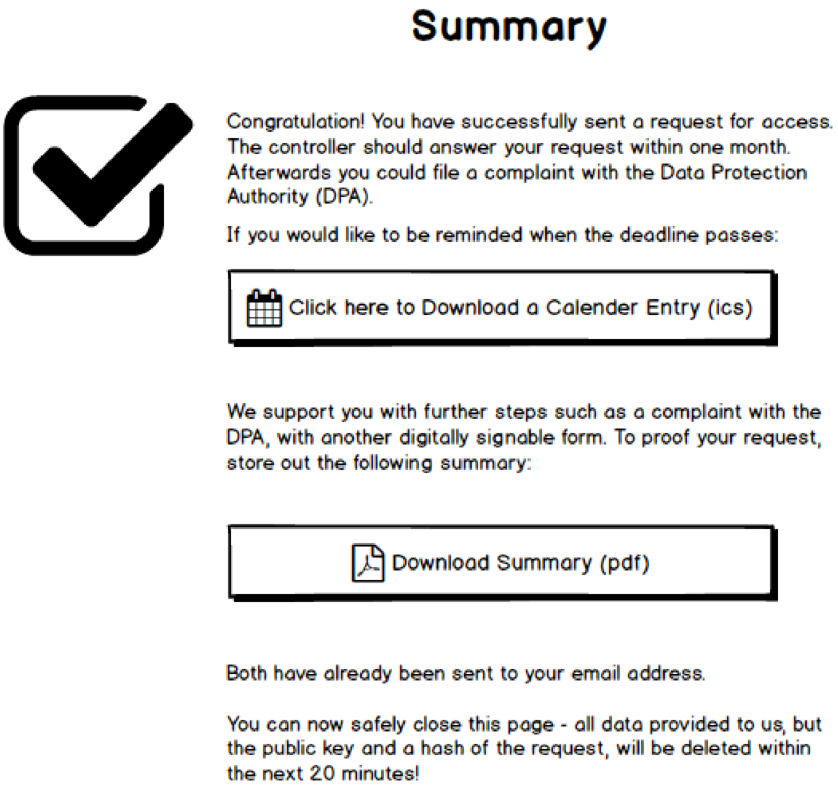
You may use this link to send the data to me securely:
https://tbd.com/?
mail=dominik.schmelz@dataprotector.at&pub=45883ee772df0f
9c51ec6233e829239d9f413e1d43c24bc8efb108e1c51c8553

Yours sincerely,
Dominik Schmelz

If want to send the above email to the controller and put a fingerprint of the message on a public blockchain to proof the request, then click send.

You will receive a summary of the request afterwards.
Click here to see the next steps and what is being done with your data.

The data subject receives a success message and confirmation via email if everything goes well. This message and the email include a summary of the request as a PDF and a calendar entry (ICS file) to be stored in the data subjects calendar software. The calendar entry reminds the data subject of the date it can file a complaint with the data protection authority if the controller has not answered the request. In case of a delay, the data subject must receive a notification from the controller; in this case, the data subject must move the calendar entry manually. It includes a link to a form to fill out the complaint information and send it again digitally signed to the Data Protection Authority.



Summary

Congratulation! You have successfully sent a request for access. The controller should answer your request within one month. Afterwards you could file a complaint with the Data Protection Authority (DPA).

If you would like to be reminded when the deadline passes:

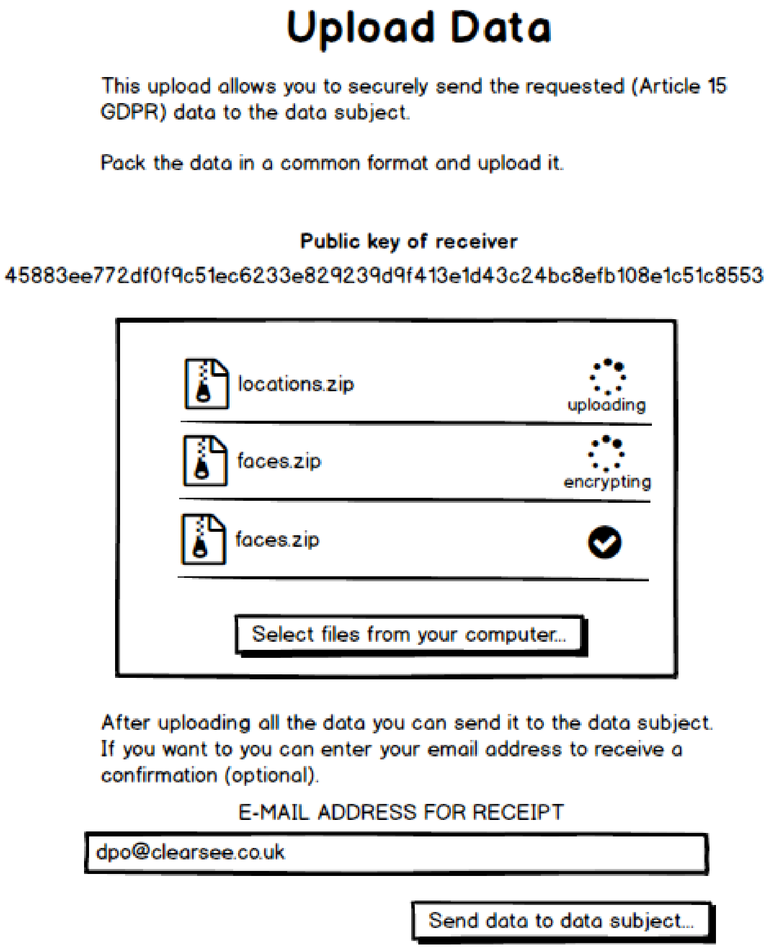
We support you with further steps such as a complaint with the DPA, with another digitally signable form. To proof your request, store out the following summary:

Both have already been sent to your email address.

You can now safely close this page - all data provided to us, but the public key and a hash of the request, will be deleted within the next 20 minutes!

The controller can open the link in its Browser and instantly gets the option to upload data. By either dragging and dropping files or selecting files with the file, browser files can be added. These are automatically encrypted. Technical Details: The web link contains the fingerprint of the public key. This is used to retrieve the public key that is then used by the Browser to encrypt the files with hybrid encryption, meaning that the data is symmetrically encrypted, and the symmetric key is encrypted with the asymmetric public key achieving faster encryption.

Optionally the controller can enter an email address to receive a summary of the uploaded files.



Upload Data

This upload allows you to securely send the requested (Article 15 GDPR) data to the data subject.

Pack the data in a common format and upload it.

Public key of receiver

45883ee772df0f9c51ec6233e829239d9f413e1d43c24bc8efb108e1c51c8553

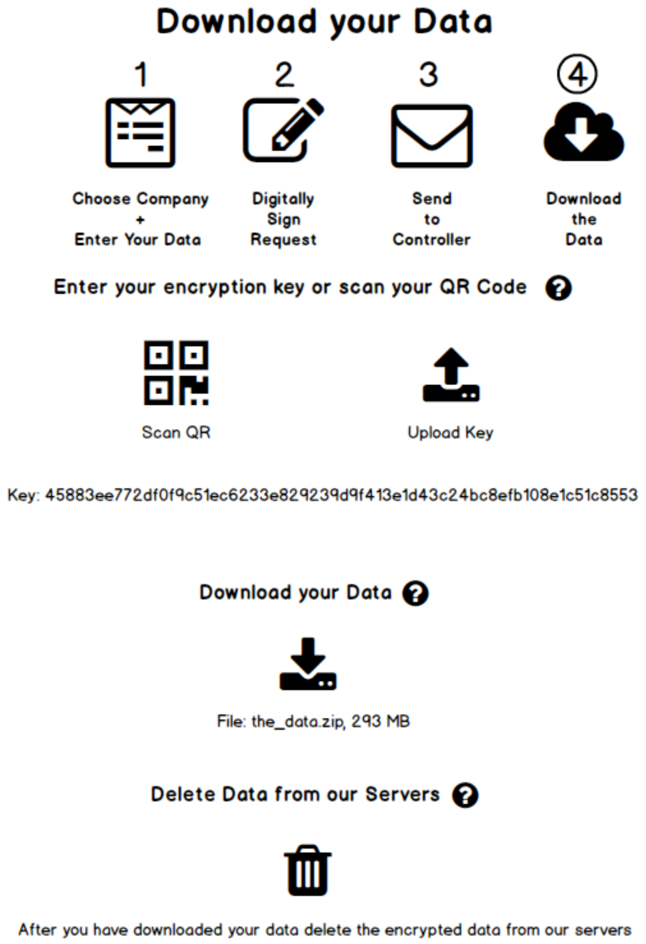
After uploading all the data you can send it to the data subject. If you want to you can enter your email address to receive a confirmation (optional).

E-MAIL ADDRESS FOR RECEIPT

dpo@clearsee.co.uk

After the data has been uploaded by the controller, the data subject receives a notification. This notification links to a webpage that allows the data subject to upload or scan its private key. Then the data is decrypted and can be downloaded. When the download has finished and the decryption is successful, a deletion request can be submitted to the server of the application. This concludes the data transfer, and no further data is stored on the application servers.

Technical Detail: The encrypted data is retrieved from the application server's binary storage; The private key is validated against the fingerprint, and the decryption is done in the browser.



Enter your encryption key or scan your QR Code ?

Scan QR

Upload Key

Key: 45883ee772df0f9c51ec6233e829239d9f413e1d43c24bc8efb108e1c51c8553

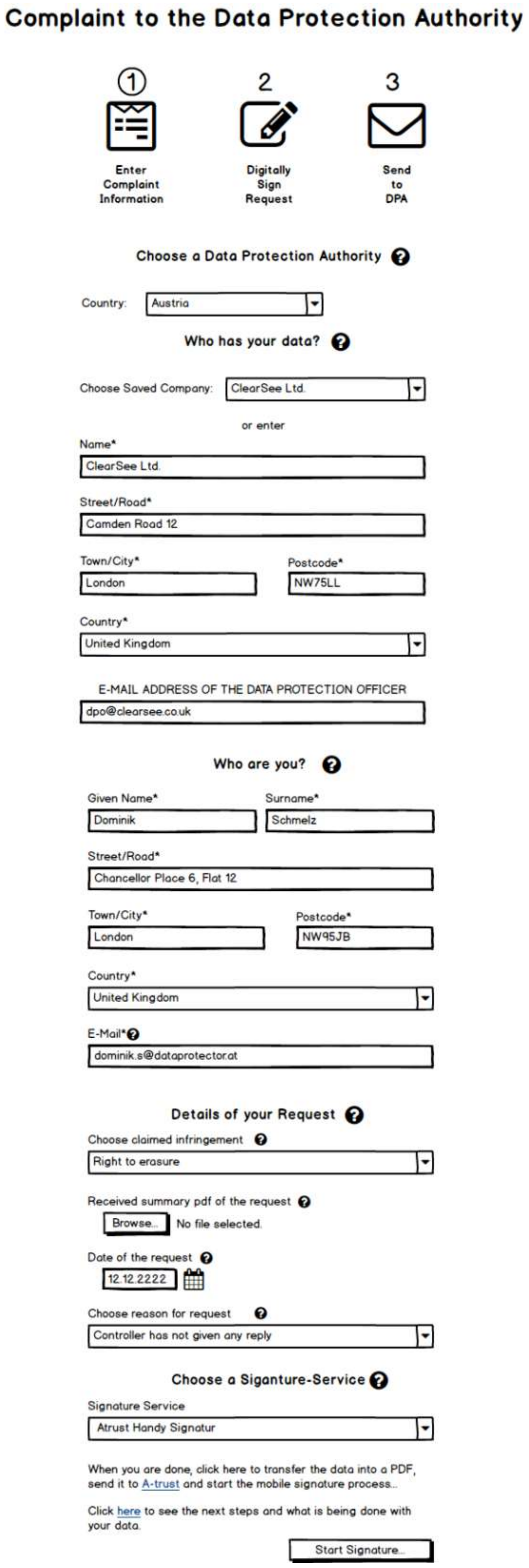
Download your Data ?

File: the_data.zip, 293 MB

Delete Data from our Servers ?

After you have downloaded your data delete the encrypted data from our servers

To file a complaint, the data subject can click the link provided in the calendar entry. The data subject can choose a data protection authority it wants to address. It must specify the details of the request. These have been derived from the Austrian data protection authorities' complaint form. The summary received via email can be attached to the complaint to prove the request to the authority. Many fields of the complaint form are automatically filled out (such as type of signature, type of contact etc.).



Complaint to the Data Protection Authority

1 Enter Complaint Information

2 Digitally Sign Request

3 Send to DPA

Choose a Data Protection Authority ?

Country: Austria ▾

Who has your data? ?

Choose Saved Company: ClearSee Ltd. ▾

or enter

Name*: ClearSee Ltd.

Street/Road*: Camden Road 12

Town/City*: London

Postcode*: NW75LL

Country*: United Kingdom ▾

E-MAIL ADDRESS OF THE DATA PROTECTION OFFICER: dpo@clearsee.co.uk

Who are you? ?

Given Name*: Dominik

Surname*: Schmelz

Street/Road*: Chancellor Place 6, Flat 12

Town/City*: London

Postcode*: NW95JB

Country*: United Kingdom ▾

E-Mail* ?: dominik.s@dataprotector.at

Details of your Request ?

Choose claimed infringement ?: Right to erasure ▾

Received summary pdf of the request ?

Browse... No file selected.

Date of the request ?: 12.12.2222

Choose reason for request ?: Controller has not given any reply ▾

Choose a Sigantur-Service ?

Signature Service: Atrust Handy Signatur ▾

When you are done, click here to transfer the data into a PDF, send it to A-trust and start the mobile signature process...

Click here to see the next steps and what is being done with your data.

Start Signature...

Now please evaluate the application by answering the following questions. If the proposed solution does not support the step (question) at all or the description does not mention a support for the step select "not supportive".

Question 12: Please answer the following criteria questions for "Create Request for Access" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.

- How well can the requested data be defined?
- How easy can the access form be found?
- How fast can a request be made?
- How well does the form check on completeness of information?
- How well does the application provide a confirmation?
- How well does the application deal with missing information?
- How well is the application accessible for disabled people?
- Is the application sufficiently multilingual?
- How trustworthy does the filing step seem?
- How well does the application implement different communication channels?

Question 13: Please answer the following criteria questions for "Check Status" S_1_2: The data subject checks whether the request has been done.

- How user friendly is the check of the status?
- How satisfactory is the frequency of status updates?
- How well does the application help to identify that a request takes longer than legally allowed?
- How fast is the status updated?
- How well is the process described and the next step predictable?
- How long does it take to log in?
- How well does the application provide information since when the request was filed?
- How trustworthy does the status seem?

Question 14: Please answer the following criteria questions for "Receive Result" S_1_3: The data subject receives a response from the data controller.

- How well can the response be understood?
- How secure is the data transfer?
- How easy is the retrieval of the result?
- How well does the application check wether the response is in the legally allowed response time?
- How well does the application support consistency of the response?
- Are there sufficient ways to receive the result notification?

Question 15: Please answer the following criteria questions for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.

- How simple is the identification process?
- How secure is the identification for the user?
- How strong is the proof of identity in the means of reasonable doubts?
- How effortless are the prerequisites needed for the identification?
- How trustworthy does the identification seem?
- How long does it take to check the validity of the identification?
- Is the identification sufficiently localized for different countries?

Question 16: Please answer the following criteria questions for "Check Delay Explanation" S_6: The data subject receives and checks a result stating that only parts of its data have been accessed.

- How well does the application explain the delay?
- How fast can the delay explanation be accessed?
- How well does the application show the remaining process?
- How well does the application help by moving forward in case of a delay?
- How well does the application support communication channels for receiving the delay notification?
- How well does the communication with a processor work?

Question 17: Please answer the following criteria questions for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.

- How fast can the status of waiting for response be accessed?
- How predictable is the further coming process when waiting for a response?
- How well does the application support the user with filing a complaint after the legally required response time?

Question 18: Please answer the following criteria questions for "Check Rejection" S_8: If the response was a rejection, the data subject checks the reason for the rejection.

- How well does the software allow for a complaint in case of a rejection?
- How well does the application support understanding the rejection?
- How trustworthy is the step of rejection?
- How well is the process traceable?

Question 19: Please answer the following criteria questions for "Open and Understand Received Data" S_9: After receiving the data files, the data subject tries to open and understand the data.

- How fast can the data be accessed?
- How well does the application allow for an export of the data?
- How well does the application help checking that the format is an industry standard?
- How well does the application support understanding the data?

Question 20: Please answer the following criteria questions for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.

- How well does the system explain the complaint process and its requirements sufficiently?
- How fault tolerant is the form for filing a complaint?
- How well does the system allow for a history export (audit trail)?
- How well is trust in the system for filing a complaint assured?
- How well does the system provide options for further appeals?
- How sufficient is the selection of languages to file the complaint in?

Question 21: Please answer the following criteria questions for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.

- How user-friendly is the process of receiving a result of a complaint?
- How well does the system give further remedies against a decision?
- How sufficient is the transparency of the process of a complaint?
- How well is trust in the system for filing a complaint assured?

E.1.6 Results of Questionnaire of the Expert Interview

Table E.1: Results of question 2 of the expert interview

<table>
<tr><td colspan="13">Q2. Please answer the following criteria questions for "Create Request for Erasure" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.</td></tr>
<tr><td></td><td colspan="2">Not supportive</td><td colspan="2">Little supportive</td><td colspan="2">Supportive</td><td colspan="2">Well supportive</td><td colspan="2">Very supportive</td><td>Total</td><td>Weighted Average</td></tr>
<tr><td>How easy can the deletion form be found?</td><td>14,29%</td><td>1</td><td>0,00%</td><td>0</td><td>28,57%</td><td>2</td><td>28,57%</td><td>2</td><td>28,57%</td><td>2</td><td>7</td><td>3,57</td></tr>
</table>

How much time is needed to file the request?	0,00%	0	14,29%	1	0,00%	0	28,57%	2	57,14%	4	7	4,29
How well can the data to be deleted be defined?	0,00%	0	14,29%	1	14,29%	1	28,57%	2	42,86%	3	7	4
How well is the process for creating request for erasure explained?	0,00%	0	0,00%	0	14,29%	1	42,86%	3	42,86%	3	7	4,29
How well is the use of templates when creating a request?	0,00%	0	0,00%	0	28,57%	2	28,57%	2	42,86%	3	7	4,14
How well does the form check on completeness of information when creating a request?	0,00%	0	0,00%	0	14,29%	1	0,00%	0	85,71%	6	7	4,71

Table E.2: Results of question 3 of the expert interview

Q3. Please answer the following criteria questions for "Check Status" S_1_2: The data subject checks whether the request has been done.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How simple is it to check the status of the request?	0,00%	0	14,29%	1	0,00%	0	28,57%	2	57,14%	4	7	4,29
How transparent are the system and status updates to a user?	14,29%	1	0,00%	0	42,86%	3	14,29%	1	28,57%	2	7	3,43
How understandable is the status presented?	14,29%	1	0,00%	0	14,29%	1	42,86%	3	28,57%	2	7	3,71

How predictable is the process and status updates?	0,00%	0	14,29%	1	28,57%	2	42,86%	3	14,29%	1	7	3,57
How satisfactory is the frequency of the updates?	28,57%	2	14,29%	1	14,29%	1	42,86%	3	0,00%	0	7	2,71
Are the updates available over sufficient communication channels?	14,29%	1	0,00%	0	28,57%	2	42,86%	3	14,29%	1	7	3,43
How well is the availability of the updates guaranteed?	14,29%	1	0,00%	0	42,86%	3	0,00%	0	42,86%	3	7	3,57

Table E.3: Results of question 4 of the expert interview

Q4. Please answer the following criteria questions for "Receive Result" S_1_3: The data subject receives a response from the data controller.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How understandable are the results provided?	14,29%	1	0,00%	0	0,00%	0	14,29%	1	71,43%	5	7	4,29
How usable are the results provided?	14,29%	1	0,00%	0	0,00%	0	14,29%	1	71,43%	5	7	4,29
How private are the results provided?	14,29%	1	0,00%	0	14,29%	1	28,57%	2	42,86%	3	7	3,86
How satisfactory is the user notification about a result?	14,29%	1	0,00%	0	42,86%	3	0,00%	0	42,86%	3	7	3,57
How well does the system provide a way for the controller to contact the subject?	14,29%	1	0,00%	0	14,29%	1	14,29%	1	57,14%	4	7	4

How well does the system provide a timely response?	14,29%	1	0,00%	0	28,57%	2	14,29%	1	42,86%	3	7	3,71
How well does the system guarantee consistency of the results?	14,29%	1	0,00%	0	28,57%	2	28,57%	2	28,57%	2	7	3,57
How satisfactory are different communication channels supported?	14,29%	1	0,00%	0	28,57%	2	42,86%	3	14,29%	1	7	3,43

Table E.4: Results of question 5 of the expert interview

Q5. Please answer the following criteria questions for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How secure is the proof of identity for the user?	0,00%	0	0,00%	0	0,00%	0	14,29%	1	85,71%	6	7	4,86
How strong is the proof of identity in the means of reasonable doubts?	0,00%	0	0,00%	0	0,00%	0	28,57%	2	71,43%	5	7	4,71
How hard are are the prerequisites to achieve?	0,00%	0	0,00%	0	0,00%	0	57,14%	4	42,86%	3	7	4,43
How easy can the identity be proven?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43
How satisfactory are the options to proof the identity?	0,00%	0	0,00%	0	42,86%	3	14,29%	1	42,86%	3	7	4
How well can the identity be proven on multiple devices?	0,00%	0	14,29%	1	14,29%	1	14,29%	1	57,14%	4	7	4,14

Table E.5: Results of question 6 of the expert interview

Q6. Please answer the following criteria questions for "Check Delay Explanation" S_6: The data subject receives and checks a notification of a delayed result with an explanation to the data subject.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How simple and clear is the delay explanation?	14,29%	1	14,29%	1	14,29%	1	0,00%	0	57,14%	4	7	3,71
How well are the further rights of the data subject supported?	14,29%	1	14,29%	1	14,29%	1	0,00%	0	57,14%	4	7	3,71
How simple is it to retrieve the explanation?	14,29%	1	0,00%	0	14,29%	1	14,29%	1	57,14%	4	7	4
How fast is the delay sent after the delay is obvious?	14,29%	1	0,00%	0	14,29%	1	57,14%	4	14,29%	1	7	3,57
How satisfactory are the communication channels supported?	14,29%	1	0,00%	0	42,86%	3	42,86%	3	0,00%	0	7	3,14
How well does the system provide a way for the subject to contact a processor?	14,29%	1	0,00%	0	28,57%	2	14,29%	1	42,86%	3	7	3,71
How well is the consistency of the delay explanation guaranteed?	14,29%	1	0,00%	0	14,29%	1	28,57%	2	42,86%	3	7	3,86
How satisfactory can the delay be confirmed?	14,29%	1	0,00%	0	28,57%	2	28,57%	2	28,57%	2	7	3,57

Table E.6: Results of question 7 of the expert interview

Q7. Please answer the following criteria questions for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How satisfactory does the system ensure that the subject knows the legal response time?	0,00%	0	0,00%	0	42,86%	3	14,29%	1	42,86%	3	7	4
How satisfactory does the system support the subject with creating a complaint?	0,00%	0	0,00%	0	14,29%	1	14,29%	1	71,43%	5	7	4,57
How fast is the delay message received?	14,29%	1	0,00%	0	28,57%	2	28,57%	2	28,57%	2	7	3,57
How satisfactory is the status and next steps presented to the data subject?	0,00%	0	0,00%	0	42,86%	3	14,29%	1	42,86%	3	7	4
How satisfactory are the options for the controller to automate the process?	14,29%	1	28,57%	2	14,29%	1	14,29%	1	28,57%	2	7	3,14

Table E.7: Results of question 8 of the expert interview

Q8. Please answer the following criteria questions for "Check Positive Result" S_8: The data subject receives and checks a positive result (data erased).		

	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How well does the system ensure that the result has been within the legally required response time?	14,29%	1	0,00%	0	28,57%	2	0,00%	0	57,14%	4	7	3,86
How well does the system provide information on what has been deleted in a clear manner?	14,29%	1	0,00%	0	14,29%	1	28,57%	2	42,86%	3	7	3,86
How well does the system check that all necessary information has been provided?	14,29%	1	0,00%	0	28,57%	2	28,57%	2	28,57%	2	7	3,57
How satisfactory are the ways to save the results and history?	14,29%	1	14,29%	1	28,57%	2	14,29%	1	28,57%	2	7	3,29
How convenient is the check of the positive result?	14,29%	1	0,00%	0	0,00%	0	28,57%	2	57,14%	4	7	4,14

Table E.8: Results of question 9 of the expert interview

Q9. Please answer the following criteria questions for "Check Partial Erasure Explanation" S_9: The data subject receives and checks a result stating that only parts of its data have been deleted.							
	Not supportive	Little supportive	Supportive	Well supportive	Very supportive	Total	Weighted Average

How well does the system provide options for a complaint?	0,00%	0	14,29%	1	0,00%	0	14,29%	1	71,43%	5	7	4,43
How well does the system explain all terminology in a sufficient way?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43
How well does the system provide information on what has been and what has not been deleted in a clear manner?	14,29%	1	0,00%	0	14,29%	1	28,57%	2	42,86%	3	7	3,86

Table E.9: Results of question 10 of the expert interview

Q10. Please answer the following criteria questions for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How easy is it to file a complaint?	0,00%	0	0,00%	0	0,00%	0	28,57%	2	71,43%	5	7	4,71
How well does the system verify the complaint?	0,00%	0	0,00%	0	0,00%	0	42,86%	3	57,14%	4	7	4,57
How well does the system automate parts of the complaint process?	0,00%	0	0,00%	0	0,00%	0	57,14%	4	42,86%	3	7	4,43
How well does the system explain the complaint process sufficiently?	0,00%	0	16,67%	1	16,67%	1	16,67%	1	50,00%	3	6	4
How well is trust in the system for filing a complaint assured?	0,00%	0	0,00%	0	28,57%	2	14,29%	1	57,14%	4	7	4,29

How well can the history of the process be viewed?	0,00%	0	28,57%	2	14,29%	1	14,29%	1	42,86%	3	7	3,71

Table E.10: Results of question 11 of the expert interview

Q11. Please answer the following criteria questions for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How well does the system explain the complaint process sufficiently?	14,29%	1	0,00%	0	14,29%	1	14,29%	1	57,14%	4	7	4
How well does the system explain the complaint result?	14,29%	1	0,00%	0	0,00%	0	28,57%	2	57,14%	4	7	4,14
How well does the system provide options for further appeals?	14,29%	1	0,00%	0	0,00%	0	14,29%	1	71,43%	5	7	4,29

Table E.11: Results of question 12 of the expert interview

Q12. Please answer the following criteria questions for "Create Request for Access" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How well can the requested data be defined?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43
How easy can the access form be found?	0,00%	0	14,29%	1	0,00%	0	28,57%	2	57,14%	4	7	4,29

How fast can a request be made?	0,00%	0	0,00%	0	0,00%	0	42,86%	3	57,14%	4	7	4,57
How well does the form check on completeness of information?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43
How well does the application provide a confirmation?	0,00%	0	0,00%	0	0,00%	0	28,57%	2	71,43%	5	7	4,71
How well does the application deal with missing information?	0,00%	0	0,00%	0	0,00%	0	42,86%	3	57,14%	4	7	4,57
How well is the application accessible for disabled people?	14,29%	1	14,29%	1	28,57%	2	0,00%	0	42,86%	3	7	3,43
Is the application sufficiently multilingual?	42,86%	3	0,00%	0	0,00%	0	14,29%	1	42,86%	3	7	3,14
How trustworthy does the filing step seem?	0,00%	0	0,00%	0	28,57%	2	28,57%	2	42,86%	3	7	4,14
How well does the application implement different communication channels?	28,57%	2	0,00%	0	14,29%	1	57,14%	4	0,00%	0	7	3

Table E.12: Results of question 13 of the expert interview

Q13. Please answer the following criteria questions for "Check Status" S_1_2: The data subject checks whether the request has been done.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How user friendly is the check of the status?	0,00%	0	14,29%	1	14,29%	1	28,57%	2	42,86%	3	7	4

How satisfactory is the frequency of status updates?	14,29%	1	0,00%	0	42,86%	3	0,00%	0	42,86%	3	7	3,57
How well does the application help to identify that a request takes longer than legally allowed?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43
How fast is the status updated?	14,29%	1	0,00%	0	42,86%	3	14,29%	1	28,57%	2	7	3,43
How well is the process described and the next step predictable?	0,00%	0	14,29%	1	14,29%	1	14,29%	1	57,14%	4	7	4,14
How long does it take to log in?	14,29%	1	0,00%	0	28,57%	2	28,57%	2	28,57%	2	7	3,57
How well does the application provide information since when the request was filed?	0,00%	0	14,29%	1	42,86%	3	14,29%	1	28,57%	2	7	3,57
How trustworthy does the status seem?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43

Table E.13: Results of question 14 of the expert interview

Q14. Please answer the following criteria questions for "Receive Result" S_1_3: The data subject receives a response from the data controller.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How well can the response be understood?	14,29%	1	0,00%	0	14,29%	1	14,29%	1	57,14%	4	7	4
How secure is the data transfer?	0,00%	0	0,00%	0	28,57%	2	14,29%	1	57,14%	4	7	4,29

How easy is the retrieval of the result?	0,00%	0	0,00%	0	42,86%	3	14,29%	1	42,86%	3	7	4
How well does the application check wether the response is in the legally allowed response time?	28,57%	2	0,00%	0	28,57%	2	14,29%	1	28,57%	2	7	3,14
How well does the application support consistency of the response?	14,29%	1	14,29%	1	0,00%	0	42,86%	3	28,57%	2	7	3,57
Are there sufficient ways to receive the result notification?	14,29%	1	0,00%	0	42,86%	3	14,29%	1	28,57%	2	7	3,43

Table E.14: Results of question 15 of the expert interview

Q15. Please answer the following criteria questions for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How simple is the identification process?	0,00%	0	14,29%	1	14,29%	1	14,29%	1	57,14%	4	7	4,14
How secure is the identification for the user?	0,00%	0	14,29%	1	0,00%	0	14,29%	1	71,43%	5	7	4,43
How strong is the proof of identity in the means of reasonable doubts?	14,29%	1	0,00%	0	0,00%	0	28,57%	2	57,14%	4	7	4,14
How effortless are the prerequisites needed for the identification?	0,00%	0	14,29%	1	28,57%	2	28,57%	2	28,57%	2	7	3,71

How trustworthy does the identification seem?	0,00%	0	0,00%	0	14,29%	1	14,29%	1	71,43%	5	7	4,57
How long does it take to check the validity of the identification?	14,29%	1	0,00%	0	0,00%	0	28,57%	2	57,14%	4	7	4,14
Is the identification sufficiently localized for different countries?	14,29%	1	14,29%	1	28,57%	2	14,29%	1	28,57%	2	7	3,29

Table E.15: Results of question 16 of the expert interview

Q16. Please answer the following criteria questions for "Check Delay Explanation" S_6: The data subject receives and checks a result stating that only parts of its data have been accessed.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How well does the application explain the delay?	28,57%	2	0,00%	0	0,00%	0	28,57%	2	42,86%	3	7	3,57
How fast can the delay explanation be accessed?	28,57%	2	0,00%	0	42,86%	3	14,29%	1	14,29%	1	7	2,86
How well does the application show the remaining process?	28,57%	2	0,00%	0	14,29%	1	14,29%	1	42,86%	3	7	3,43
How well does the application help by moving forward in case of a delay?	14,29%	1	14,29%	1	14,29%	1	28,57%	2	28,57%	2	7	3,43
How well does the application support communication channels for receiving the delay notification?	28,57%	2	0,00%	0	14,29%	1	42,86%	3	14,29%	1	7	3,14

How well does the communication with a processor work?	28,57%	2	0,00%	0	14,29%	1	14,29%	1	42,86%	3	7	3,43

Table E.16: Results of question 17 of the expert interview

Q17. Please answer the following criteria questions for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How fast can the status of waiting for response be accessed?	0,00%	0	28,57%	2	28,57%	2	0,00%	0	42,86%	3	7	3,57
How predictable is the further coming process when waiting for a response?	0,00%	0	0,00%	0	28,57%	2	28,57%	2	42,86%	3	7	4,14
How well does the application support the user with filing a complaint after the legally required response time?	0,00%	0	14,29%	1	0,00%	0	28,57%	2	57,14%	4	7	4,29

Table E.17: Results of question 18 of the expert interview

Q18. Please answer the following criteria questions for "Check Rejection" S_8: If the response was a rejection, the data subject checks the reason for the rejection.							
	Not supportive	Little supportive	Supportive	Well supportive	Very supportive	Total	Weighted Average

How well does the software allow for a complaint in case of a rejection?	28,57%	2	0,00%	0	14,29%	1	0,00%	0	57,14%	4	7	3,57
How well does the application support understanding the rejection?	28,57%	2	0,00%	0	14,29%	1	14,29%	1	42,86%	3	7	3,43
How trustworthy is the step of rejection?	28,57%	2	0,00%	0	0,00%	0	14,29%	1	57,14%	4	7	3,71
How well is the process traceable?	28,57%	2	0,00%	0	14,29%	1	14,29%	1	42,86%	3	7	3,43

Table E.18: Results of question 19 of the expert interview

Q19. Please answer the following criteria questions for "Open and Understand Received Data" S_9: After receiving the data files, the data subject tries to open and understand the data.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How fast can the data be accessed?	0,00%	0	14,29%	1	57,14%	4	0,00%	0	28,57%	2	7	3,43
How well does the application allow for an export of the data?	14,29%	1	0,00%	0	14,29%	1	28,57%	2	42,86%	3	7	3,86
How well does the application help checking that the format is an industry standard?	28,57%	2	0,00%	0	14,29%	1	14,29%	1	42,86%	3	7	3,43
How well does the application support understanding the data?	28,57%	2	0,00%	0	28,57%	2	14,29%	1	28,57%	2	7	3,14

Table E.19: Results of question 20 of the expert interview

Q20. Please answer the following criteria questions for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.												
	Not supportive		Little supportive		Supportive		Well supportive		Very supportive		Total	Weighted Average
How well does the system explain the complaint process and its requirements sufficiently?	0,00%	0	0,00%	0	14,29%	1	28,57%	2	57,14%	4	7	4,43
How fault tolerant is the form for filing a complaint?	14,29%	1	14,29%	1	0,00%	0	14,29%	1	57,14%	4	7	3,86
How well does the system allow for a history export (audit trail)?	28,57%	2	0,00%	0	28,57%	2	0,00%	0	42,86%	3	7	3,29
How well is trust in the system for filing a complaint assured?	14,29%	1	0,00%	0	28,57%	2	14,29%	1	42,86%	3	7	3,71
How well does the system provide options for further appeals?	14,29%	1	14,29%	1	0,00%	0	28,57%	2	42,86%	3	7	3,71
How sufficient is the selection of languages to file the complaint in?	28,57%	2	0,00%	0	57,14%	4	0,00%	0	14,29%	1	7	2,71

Table E.20: Results of question 21 of the expert interview

Q21. Please answer the following criteria questions for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.							
	Not supportive	Little supportive	Supportive	Well supportive	Very supportive	Total	Weighted Average

How user-friendly is the process of receiving a result of a complaint?	14,29%	1	0,00%	0	28,57%	2	14,29%	1	42,86%	3	7	3,71
How well does the system give further remedies against a decision?	28,57%	2	0,00%	0	28,57%	2	0,00%	0	42,86%	3	7	3,29
How sufficient is the transparency of the process of a complaint?	14,29%	1	14,29%	1	14,29%	1	14,29%	1	42,86%	3	7	3,57
How well is trust in the system for filing a complaint assured?	14,29%	1	14,29%	1	0,00%	0	14,29%	1	57,14%	4	7	3,86

E.1.7 Analysis of Questionnaire of the Expert Interview

Table E.21: Analysis of question 2 of the expert interview

Q2. Please answer the following criteria questions for "Create Request for Erasure" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for erasure.		
Question Identification	Application Support Score	Question Importance Score
RFE.S_1_1.Q_2_1	3,57	5,7
RFE.S_1_1.Q_2_2	4,29	5,1
RFE.S_1_1.Q_2_3	4	3,7
RFE.S_1_1.Q_2_4	4,29	3,2
RFE.S_1_1.Q_2_5	4,14	1,8
RFE.S_1_1.Q_2_6	4,71	1,5
	Weighted Score	4,1

Table E.22: Analysis of question 3 of the expert interview

Q3. Please answer the following criteria questions for "Check Status" S_1_2: The data subject checks whether the request has been done.		
Question Identification	Application Support Score	Question Importance Score
RFE.S_1_2.Q_2_1	4,29	6,9
RFE.S_1_2.Q_2_2	3,43	6
RFE.S_1_2.Q_2_3	3,71	5,1

RFE.S_1_2.Q_2_4	3,57	3,9
RFE.S_1_2.Q_2_5	2,71	3,1
RFE.S_1_2.Q_2_6	3,43	1,9
RFE.S_1_2.Q_2_7	3,57	1,1
	Weighted Score	3,6

Table E.23: Analysis of question 4 of the expert interview

Q4. Please answer the following criteria questions for "Receive Result" S_1_3: The data subject receives a response from the data controller.		
Question Identification	Application Support Score	Question Importance Score
RFE.S_1_3.Q_2_1	4,29	7,8
RFE.S_1_3.Q_2_2	4,29	7
RFE.S_1_3.Q_2_3	3,86	5,8
RFE.S_1_3.Q_2_4	3,57	5,3
RFE.S_1_3.Q_2_5	4	4,1
RFE.S_1_3.Q_2_6	3,71	2,9
RFE.S_1_3.Q_2_7	3,57	1,9
RFE.S_1_3.Q_2_8	3,43	1,2
	Weighted Score	4,0

Table E.24: Analysis of question 5 of the expert interview

Q5. Please answer the following criteria questions for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.		
Question Identification	Application Support Score	Question Importance Score
RFE.S_2.Q_2_1	4,86	5,5
RFE.S_2.Q_2_2	4,71	4,7
RFE.S_2.Q_2_3	4,43	3,5
RFE.S_2.Q_2_4	4,43	3,5
RFE.S_2.Q_2_5	4	2,2
RFE.S_2.Q_2_6	4,14	1,6
	Weighted Score	4,5

Table E.25: Analysis of question 6 of the expert interview

Q6. Please answer the following criteria questions for "Check Delay Explanation" S_6: The data subject receives and checks a notification of a delayed result with an explanation to the data subject.		
Question Identification	Application Support Score	Question Importance Score

RFE.S_6.Q_2_1	3,71	7,9
RFE.S_6.Q_2_2	3,71	6,6
RFE.S_6.Q_2_3	4	5,9
RFE.S_6.Q_2_4	3,57	4,9
RFE.S_6.Q_2_5	3,14	4,4
RFE.S_6.Q_2_6	3,71	2,9
RFE.S_6.Q_2_7	3,86	2,4
RFE.S_6.Q_2_8	3,57	1
	Weighted Score	3,7

Table E.26: Analysis of question 7 of the expert interview

Q7. Please answer the following criteria questions for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.		
Question Identification	Application Support Score	Question Importance Score
RFE.S_7.Q_2_1	4	4,9
RFE.S_7.Q_2_2	4,57	4
RFE.S_7.Q_2_3	3,57	3,1
RFE.S_7.Q_2_4	4	1,9
RFE.S_7.Q_2_5	3,14	1,1
	Weighted Score	4,0

Table E.27: Analysis of question 8 of the expert interview

Q8. Please answer the following criteria questions for "Check Positive Result" S_8: The data subject receives and checks a positive result (data erased).		
Question Identification	Application Support Score	Question Importance Score
RFE.S_8.Q_2_1	3,86	4,6
RFE.S_8.Q_2_2	3,86	4
RFE.S_8.Q_2_3	3,57	2,8
RFE.S_8.Q_2_4	3,29	1,8
RFE.S_8.Q_2_5	4,14	1,8
	Weighted Score	3,8

Table E.28: Analysis of question 9 of the expert interview

<table>
<tr><td colspan="3">Q9. Please answer the following criteria questions for "Check Partial Erasure Explanation" S_9: The data subject receives and checks a result stating that only parts of its data have been deleted.</td></tr>
<tr><td>Question Identification</td><td>Application Support Score</td><td>Question Importance Score</td></tr>
<tr><td>RFE.S_9.Q_2_1</td><td>4,43</td><td>2,9</td></tr>
<tr><td>RFE.S_9.Q_2_2</td><td>4,43</td><td>1,8</td></tr>
<tr><td>RFE.S_9.Q_2_3</td><td>3,86</td><td>1,3</td></tr>
<tr><td></td><td>Weighted Score</td><td>4,3</td></tr>
</table>

Table E.29: Analysis of question 10 of the expert interview

<table>
<tr><td colspan="3">Q10. Please answer the following criteria questions for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.</td></tr>
<tr><td>Question Identification</td><td>Application Support Score</td><td>Question Importance Score</td></tr>
<tr><td>RFE.S_4.Q_2_1</td><td>4,71</td><td>6</td></tr>
<tr><td>RFE.S_4.Q_2_2</td><td>4,57</td><td>4,8</td></tr>
<tr><td>RFE.S_4.Q_2_3</td><td>4,43</td><td>3,7</td></tr>
<tr><td>RFE.S_4.Q_2_4</td><td>4</td><td>2,8</td></tr>
<tr><td>RFE.S_4.Q_2_5</td><td>4,29</td><td>1,9</td></tr>
<tr><td>RFE.S_4.Q_2_6</td><td>3,71</td><td>1,8</td></tr>
<tr><td></td><td>Weighted Score</td><td>4,4</td></tr>
</table>

Table E.30: Analysis of question 11 of the expert interview

<table>
<tr><td colspan="3">Q11. Please answer the following criteria questions for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.</td></tr>
<tr><td>Question Identification</td><td>Application Support Score</td><td>Question Importance Score</td></tr>
<tr><td>RFE.S_5.Q_2_1</td><td>4</td><td>2,9</td></tr>
<tr><td>RFE.S_5.Q_2_2</td><td>4,14</td><td>2</td></tr>
<tr><td>RFE.S_5.Q_2_3</td><td>4,29</td><td>1,1</td></tr>
<tr><td></td><td>Weighted Score</td><td>4,1</td></tr>
</table>

Table E.31: Analysis of question 12 of the expert interview

<table>
<tr><td colspan="3">Q12. Please answer the following criteria questions for "Create Request for Access" S_1_1: The data subject finds a way of communication with the data controller (online, phone, website etc.) and files a request for access to its data.</td></tr>
<tr><td>Question Identification</td><td>Application Support Score</td><td>Question Importance Score</td></tr>
<tr><td>RFA.S_1_1.Q_2_1</td><td>4,43</td><td>9,4</td></tr>
</table>

RFA.S_1_1.Q_2_2	4,29	9
RFA.S_1_1.Q_2_3	4,57	8,4
RFA.S_1_1.Q_2_4	4,43	6,7
RFA.S_1_1.Q_2_5	4,71	5,5
RFA.S_1_1.Q_2_6	4,57	4,8
RFA.S_1_1.Q_2_7	3,43	3,8
RFA.S_1_1.Q_2_8	3,14	2,6
RFA.S_1_1.Q_2_9	4,14	3,2
RFA.S_1_1.Q_2_10	3	1,6
	Weighted Score	4,3

Table E.32: Analysis of question 13 of the expert interview

Q13. Please answer the following criteria questions for "Check Status" S_1_2: The data subject checks whether the request has been done.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_1_2.Q_2_1	4	7,8
RFA.S_1_2.Q_2_2	3,57	6,7
RFA.S_1_2.Q_2_3	4,43	5,7
RFA.S_1_2.Q_2_4	3,43	5,3
RFA.S_1_2.Q_2_5	4,14	3,8
RFA.S_1_2.Q_2_6	3,57	3,4
RFA.S_1_2.Q_2_7	3,57	1,8
RFA.S_1_2.Q_2_8	4,43	1,5
	Weighted Score	3,9

Table E.33: Analysis of question 14 of the expert interview

Q14. Please answer the following criteria questions for "Receive Result" S_1_3: The data subject receives a response from the data controller.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_1_3.Q_2_1	4	5,7
RFA.S_1_3.Q_2_2	4,29	4,8
RFA.S_1_3.Q_2_3	4	4,5
RFA.S_1_3.Q_2_4	3,14	2,8
RFA.S_1_3.Q_2_5	3,57	2
RFA.S_1_3.Q_2_6	3,43	1,2
	Weighted Score	3,9

Table E.34: Analysis of question 15 of the expert interview

Q15. Please answer the following criteria questions for "Proof Identity" S_2: The data subject proofs its identity to the data controller on request.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_2.Q_2_1	4,14	6,9
RFA.S_2.Q_2_2	4,43	5,9
RFA.S_2.Q_2_3	4,14	4,9
RFA.S_2.Q_2_4	3,71	4,1
RFA.S_2.Q_2_5	4,57	3
RFA.S_2.Q_2_6	4,14	2,2
RFA.S_2.Q_2_7	3,29	1
	Weighted Score	4,2

Table E.35: Analysis of question 16 of the expert interview

Q16. Please answer the following criteria questions for "Check Delay Explanation" S_6: The data subject receives and checks a result stating that only parts of its data have been accessed.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_6.Q_2_1	3,57	6
RFA.S_6.Q_2_2	2,86	5
RFA.S_6.Q_2_3	3,43	4
RFA.S_6.Q_2_4	3,43	2,9
RFA.S_6.Q_2_5	3,14	2,1
RFA.S_6.Q_2_6	3,43	1
	Weighted Score	3,3

Table E.36: Analysis of question 17 of the expert interview

Q17. Please answer the following criteria questions for "Wait for Delayed Result" S_7: If the first response was a notification of a delayed result; the data subject waits until it gets the results. After 2 months of waiting, the data subject may file a complaint.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_7.Q_2_1	3,57	2,7
RFA.S_7.Q_2_2	4,14	2,2
RFA.S_7.Q_2_3	4,29	1,1
	Weighted Score	3,9

Table E.37: Analysis of question 18 of the expert interview

Q18. Please answer the following criteria questions for "Check Rejection" S_8: If the response was a rejection, the data subject checks the reason for the rejection.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_8.Q_2_1	3,57	3,8
RFA.S_8.Q_2_2	3,43	3,2
RFA.S_8.Q_2_3	3,71	1,9
RFA.S_8.Q_2_4	3,43	1,1
	Weighted Score	3,5

Table E.38: Analysis of question 19 of the expert interview

Q19. Please answer the following criteria questions for "Open and Understand Received Data" S_9: After receiving the data files, the data subject tries to open and understand the data.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_9.Q_2_1	3,43	3,8
RFA.S_9.Q_2_2	3,86	3
RFA.S_9.Q_2_3	3,43	1,9
RFA.S_9.Q_2_4	3,14	1,3
	Weighted Score	3,5

Table E.39: Analysis of question 20 of the expert interview

Q20. Please answer the following criteria questions for "File Complaint" S_4: The data subject files a complaint at the Data Protection Authority.		
Question Identification	Application Support Score	Question Importance Score
RFA.S_4.Q_2_1	4,43	6
RFA.S_4.Q_2_2	3,86	4,8
RFA.S_4.Q_2_3	3,29	4,1
RFA.S_4.Q_2_4	3,71	3
RFA.S_4.Q_2_5	3,71	2,1
RFA.S_4.Q_2_6	2,71	1
	Weighted Score	3,8

Table E.40: Analysis of question 21 of the expert interview

Q21. Please answer the following criteria questions for "Receive Complaint Result" S_5: The data subject receives the result of the complaint.		
Question Identification	Application Support Score	Question Importance Score

RFA.S_5.Q_2_1	3,71	3,9
RFA.S_5.Q_2_2	3,29	2,9
RFA.S_5.Q_2_3	3,57	2
RFA.S_5.Q_2_4	3,86	1,2
	Weighted Score	3,6

List of Figures

List of Tables

Acronyms

ABE Attribute-Based Encryption. 100
AML Anti Money Laundering. 118

BPM Business Process Management. 126
BPMN Business Process Model and Notation. 126–128, 131, 132, 156, 175, 185
BSI Bundesamt für Sicherheit in der Informationstechnik. 39, 111, 112

CAGR Compound Annual Growth Rate. 67
CCPA California Consumer Privacy Act. 5, 12, 15, 16
CERT Computer Emergency Response Team. 30
CFPB Consumer Financial Protection Bureau. 20
CJEU Court of Justice of the European Union. 25, 42
CKD Child Key Derivation. 59, 109
CNIL Commission nationale de l'informatique et des libertés. 37, 138
COPPA Children's Online Privacy Protection Act. 19
CSIRT Computer Security Incident Response Team. 30
CSV Comma-separated Values. 150

DAG Directed Acyclic Graph. 48
DAO Decentralized Autonomous Organization. 51
DApp Decentralized Application. 106
DID Decentralized Identifier. 68, 69, 89
DIN Deutsches Institut für Normung. 132
DLT Distributed Ledger Technology. 47, 48, 87
DNA Deoxyribonucleic acid. 16
DoS Denial of Service. 29, 30
DPA Data Protection Authority. 37, 79, 114, 135, 138, 146, 158, 159, 165, 166, 235
DPIA Data Protection Impact Assessment. 78, 79
DPO Data Protection Officer. 81
DPPA Drivers Privacy Protection Act. 19
DSG Datenschutzgesetz. 31
DSRM Design Science Research Methodology. 3, 123, 124, 130

EBSI European blockchain service infrastructure. 69

ECC Elliptic Curve Cryptography. 106

ECJ European Court of Justice. 22

EDPD European Data Protection Board. 43

EDPS European Data Protection Supervisor. 81

EHR Electronic Health Record. 70

eIDAS electronic IDentification, Authentication and trust Services. 158, 168

ENISA European Union Agency for Cybersecurity. 39, 124

EPRS European Parliament Research Service. 25, 85–87

ESSIF European self-sovereign identity framework. 69

EU European Union. 5, 21–23, 90, 91

EUROPOL European Union Agency for Law Enforcement Cooperation. 12

FACTA Fair and Accurate Credit Transactions Act. 20

FCC Federal Communications Commission. 20

FCRA Fair Credit Reporting Act. 20

FERPA Family Educational Rights and Privacy Act. 20

FHE Fully Homomorphic Encryption. 105, 106

GDPR General Data Protection Regulation. 1–6, 12, 15, 16, 18, 19, 21–28, 32–34, 36, 38–45, 76, 77, 79, 81, 83, 86–90, 100, 108, 110, 111, 120, 121, 123, 125, 126, 130–132, 135, 139–141, 154, 157, 165, 166, 175, 177, 181, 183, 231

GLBA Gramm–Leach–Bliley Act. 20

GPS Global Positioning System. 6

HA High Availability. 30

HE Homomorphic Encryption. 105

HHS Department of Health and Human Services. 20

HIPAA Health Insurance Portability and Accountability Act. 18, 20, 111

HTTPS Hypertext Transfer Protocol Secure. 158, 166

ICO Initial Coin Offering. 47, 48

ICS Internet Calendar Scheduling. 162, 171

ICT Information and Communications Technology. 39

ID Identifier. 68

IDC International Data Corporation. 67

IoT Internet of Things. 10, 13, 14, 55, 69

IP Internet Protocol. 15, 16, 88

IPO Initial Public Offering. 47

IPv6 Internet Protocol version 6. 13

ISO International Organization for Standardization. 7, 9, 28

ISP Internet Service Provider. 88

IT Information Technology. 39

KYC Know Your Customer. 84, 88, 118

LGPD Lei Geral de Proteçao de Dados. 12

NIST National Institute of Standards and Technology. 39, 107, 111, 112

OECD Organisation for Economic Co-operation and Development. 19

OLTP Online Transactional Processing. 50

PDF Portable Document Format. 135, 139, 158, 159, 161, 162, 168, 171

PDPA Personal Data Protection Act. 5, 12

PET Privacy Enhancing Technology. 11, 45, 104, 113, 123, 124, 126

PII Personally Identifiable Information. 15

PIPL Personal Information Protection Law of the People's Republic of China. 12, 21

PKI Public Key Infrastructure. 88

PoE Proof of Existence. 69

PoS Proof of Stake. 52, 53, 61, 63–65

PoW Proof of Work. 52, 53, 61–65, 70

PRNG Pseudorandom Number Generator. 101

REST Representational State Transfer. 158

SEC Securities and Exchange Commission. 20

SEPA Single Euro Payments Area. 75

SHA-3 Secure Hash Algorithm-3. 107

SMTP Simple Mail Transfer Protocol. 159, 161, 168

SOAP Simple Object Access Protocol. 158

STO Security Token Offering. 47

SWIFT Society for Worldwide Interbank Financial Telecommunication. 85

TCPA Telephone Consumer Protection Act. 20

UDHR The Universal Declaration of Human Rights. 7, 18

URL Uniform Resource Locator. 159

VIN Vehicle Identification Number. 16, 89

Bibliography

[AA13] Meg Leta Ambrose and Jef Ausloos. The Right to Be Forgotten Across the Pond. *Journal of Information Policy*, pages 1–23, 2013.

[ADB14] Tiwalade Adelola, Ray Dawson, and Firat Batmaz. Privacy and Data Protection in E-commerce in Developing Nations: Evaluation of Different Data Protection Approaches. *Regulation*, page 5, 2014.

[AE16] EL-YAHYAOUI Ahmed and Mohamed Dafir Elkettani. Fully homomorphic encryption: state of art and comparison. *International Journal of Computer Science and Information Security (IJCSIS)*, 14(4), 2016.

[AGKK19] Myrto Arapinis, Andriana Gkaniatsou, Dimitris Karakostas, and Aggelos Kiayias. A formal treatment of hardware wallets. In *International Conference on Financial Cryptography and Data Security*, pages 426–445. Springer, 2019.

[AKR$^+$13] Elli Androulaki, Ghassan O Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun. Evaluating user privacy in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 34–51. Springer, 2013.

[AMMS19] Simone Agostinelli, Fabrizio Maria Maggi, Andrea Marrella, and Francesco Sapio. Achieving GDPR compliance of BPMN process models. In *International Conference on Advanced Information Systems Engineering*, pages 10–22. Springer, 2019.

[And11] J. Andress. *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*. Syngress basics series. Elsevier Science, 2011.

[And20] A. Anderl. *Blockchain in der Rechtspraxis*. Rechtspraxis. LexisNexis, Wien, 2020.

[Ant14] A.M. Antonopoulos. *Mastering Bitcoin: Unlocking Digital Cryptocurrencies*. O'Reilly Media, 2014.

[AOSV22] Adi Akavia, Neta Oren, Boaz Sapir, and Margarita Vald. Compact storage for homomorphic encryption. *Cryptology ePrint Archive*, 2022.

[Ass49] United Nations. General Assembly. *Universal declaration of human rights*. Department of State, United States of America, 1949.

[Bar06] Susan B Barnes. A privacy paradox: Social networking in the United States. *First Monday*, 2006.

[Bar20a] Elaine Barker. Recommendation for key management (800-57). Technical report, National Institute of Standards and Technology, May 2020.

[Bar20b] Catherine Barrett. Emerging trends from the first year of EU GDPR enforcement. *Scitech Lawyer*, pages 22–35, 2020.

[BB93] Jani G Byrne and Todd Barlow. Structured brainstorming: A method for collecting user requirements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 37, pages 427–431. SAGE Publications Sage CA: Los Angeles, CA, 1993.

[BB16] Colin J Bennett and Robin M Bayley. Privacy protection in the era of 'big data': regulatory challenges and social assessments. *Exploring the Boundaries of Big Data*, 2016.

[BB17] Luca Bolognini and Camilla Bistolfi. Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation. *Computer law & security review*, pages 171–181, 2017.

[BBB20] Matthew Budman, Rupesh Bhat, and Sayanika Bordoloi. Deloitte's 2020 Global Blockchain Survey: From promise to reality. Survey, Deloitte, 2020.

[BC06] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[BCM19a] Cesare Bartolini, Antonello Calabró, and Eda Marchetti. Enhancing Business Process Modelling with Data Protection Compliance: An Ontology-based Proposal. In *ICISSP*, pages 421–428, 2019.

[BCM19b] Cesare Bartolini, Antonello Calabró, and Eda Marchetti. GDPR and business processes: An effective solution. In *Proceedings of the 2nd International Conference on Applications of Intelligent Systems*, pages 1–5, 2019.

[BD05] Dave Brunsdon and Erica Dalziell. Making organisations resilient: understanding the reality of the challenge. 2005.

[BDPVA07] Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. Sponge functions. In *ECRYPT hash workshop*, volume 2007. Citeseer, 2007.

[BF20] Saliha Irem Besik and Johann-Christoph Freytag. Managing Consent in Workflows under GDPR. In *ZEUS*, pages 18–25, 2020.

[BFM19] Manuel Blum, Paul Feldman, and Silvio Micali. Non-interactive zero-knowledge and its applications. In *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*, pages 329–349. 2019.

[BHK+20] Vitalik Buterin, Diego Hernandez, Thor Kamphefner, Khiem Pham, Zhi Qiao, Danny Ryan, Juhyeok Sin, Ying Wang, and Yan X Zhang. Combining GHOST and Casper. *arXiv preprint arXiv:2003.03052*, 2020.

[BHL+14] Federico Boccardi, Robert W Heath, Angel Lozano, Thomas L Marzetta, and Petar Popovski. Five disruptive technology directions for 5G. *IEEE Communications Magazine*, pages 74–80, 2014.

[BHLR12] Achim D Brucker, Isabelle Hang, Gero Lückemeyer, and Raj Ruparel. SecureBPMN: Modeling and enforcing access control requirements in business processes. In *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*, pages 123–126, 2012.

[BKP14] Alex Biryukov, Dmitry Khovratovich, and Ivan Pustogarov. Deanonymisation of Clients in Bitcoin P2P Network. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 15–29, 2014.

[BMS18] Stefano Bistarelli, Ivan Mercanti, and Francesco Santini. An analysis of non-standard bitcoin transactions. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, pages 93–96. IEEE, 2018.

[Boa18] European Data Protection Board. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 2018.

[Boa19] European Data Protection Board. Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, 2019.

[Boa20] European Data Protection Board. Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications. `https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/guidelines-12020-processing-personal-data-context_en`, January 2020. (Accessed on 03/27/2021).

[Bor16] Frederik J Zuiderveen Borgesius. Singling out people without knowing their names–Behavioural targeting, pseudonymous data, and the new Data Protection Regulation. *Computer Law & Security Review*, pages 256–271, 2016.

[BR17] Colin J Bennett and Charles D Raab. *The governance of privacy: Policy instruments in global perspective*. Routledge, 2017.

[Bro68] Bernice B Brown. Delphi process: a methodology used for the elicitation of opinions of experts. Technical report, Rand Corp Santa Monica CA, 1968.

[BS18] Aleksander Berentsen and Fabian Schär. A short introduction to the world of cryptocurrencies. 2018.

[BS19] Omri Ben-Shahar. Data Pollution. *Journal of Legal Analysis*, 11:104–159, 2019.

[BSI21] BSI. Kryptographische Verfahren: Empfehlungen und Schlüssellängen. Technical report, Bundesamt für Sicherheit in der Informationstechnik, January 2021.

[Bun03] Bundesgesetzblatt. Bundesgesetz zur Gewährleistung eines hohen Sicherheitsniveaus von Netz- und Informationssystemen (Netz- und Informationssystemsicherheitsgesetz – NISG), 2003. Fassung vom 14.03.2021.

[BVE$^+$03] John Borking, P. Verhaar, B.M.A. Eck, P. Siepel, G.W. Blarkom, R. Coolen, M. Uyl, J. Holleman, P. Bison, R. Veer, J. Giezen, Andrew Patrick, C. Holmes, J.C.A. Lubbe, Roy Lachman, S. Kenny, Randy Song, K. Cartrysse, J. Huizenga, and X. Zhou. *Handbook of Privacy and Privacy-Enhancing Technologies The case of Intelligent Software Agents*. November 2003.

[Byg14] Lee Andrew Bygrave. *Data privacy law: an international perspective*. Oxford University Press, 2014.

[Car18] Peter Carey. *Data protection: a practical guide to UK and EU law*. Oxford University Press, Inc., 2018.

[CBNP10] M.D. Coogan, M.Z. Brettler, C.A. Newsom, and P. Perkins. *The New Oxford Annotated Bible with Apocrypha: New Revised Standard Version*. OUP USA, 2010.

[CCP] CCPA. The California Consumer Privacy Act of 2018.

[CDK$^+$17] Jan Camenisch, David Derler, Stephan Krenn, Henrich C Pöhls, Kai Samelin, and Daniel Slamanig. Chameleon-hashes with ephemeral trapdoors. In *IACR International Workshop on Public Key Cryptography*, pages 152–182. Springer, 2017.

[CDP18] Fran Casino, Thomas K Dasaklis, and Constantinos Patsakis. A systematic literature review of blockchain-based applications: current status, classification and open issues. *Telematics and Informatics*, 2018.

[Cen96] Centers for Medicare & Medicaid Services. The Health Insurance Portability and Accountability Act of 1996 (HIPAA). `https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996`, 1996.

[Cha17] Gauthier Chassang. The impact of the EU general data protection regulation on scientific research. *ecancermedicalscience*, 2017.

[CIMM21] Lelio Campanile, Mauro Iacono, Fiammetta Marulli, and Michele Mastroianni. Designing a GDPR compliant blockchain-based IoV distributed information tracking system. *Information Processing & Management*, 58(3):102511, 2021.

[CKY18] Jason Paul Cruz, Yuichi Kaji, and Naoto Yanai. RBAC-SC: Role-based access control using smart contract. *Ieee Access*, 6:12240–12251, 2018.

[Com18] European Commission. Ethics and data protection. Technical report, EU, November 2018.

[Cor20] International Data Corporation. Blockchain Solutions Will Continue to See Robust Investments, Led by Banking and Manufacturing, According to New IDC Spending Guide. Survey, IDC, September 2020.

[Cou16a] Council of European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, 2016.

[Cou16b] Court of Justice of the European Union. Case C-582/14 (Breyer/Germany) ECLI:EU:C:2016:779, definition of 'personal data' - internet protocol addresses - storage of data by an online media services provider - national legislation not permitting the legitimate interest pursued by the controller to be taken into account, 2016.

[Cou17] Court of Justice of the European Union. Case C-434/16 (Nowak/Data Protection Commissioner) ECLI:EU:C:2017:994, Protection of individuals with regard to the processing of personal data, 2017.

[Cou18] Court of Justice of the European Union. Case C-25/17 (Jehovan todistajat) EU:C:2018:551, 2018.

[CS05] Ramnath K Chellappa and Raymond G Sin. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information technology and management*, pages 181–202, 2005.

[ct20] DLA Piper's cybersecurity team. DLA Piper GDPR Data Breach Survey 2020. Survey, DLA Piper, 2020.

[CV17] Christian Cachin and Marko Vukolić. Blockchain consensus protocols in the wild. *arXiv preprint arXiv:1707.01873*, 2017.

[CVG$^+$15] Jan Claes, Irene Vanderfeesten, Frederik Gailly, Paul Grefen, and Geert Poels. The Structured Process Modeling Theory (SPMT) a cognitive view on why and how modelers benefit from structuring the process of process modeling. *Information Systems Frontiers*, 17(6):1401–1425, 12 2015.

[CWVK11] Aaron Ciaghi, Komminist Weldemariam, Adolfo Villafiorita, and Fondazione Kessler. Law Modeling with Ontological Support and BPMN: a Case Study. 01 2011.

[D$^+$15] Morris J Dworkin et al. Sha-3 standard: Permutation-based hash and extendable-output functions. 2015.

[Dan04] George Danezis. The traffic analysis of continuous-time mixes. In *International Workshop on Privacy Enhancing Technologies*, pages 35–50. Springer, 2004.

[DB16] Nicola Döring and Jürgen Bortz. Forschungsmethoden und evaluation. *Wiesbaden: Springerverlag*, 2016.

[DDFH$^+$15] George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirtea, and Stefan Schiffner. Privacy and Data Protection by Design-from policy to engineering. 2015.

[DeC18] J.W. DeCew. *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology*. Cornell University Press, 2018.

[Den82] Dorothy Elizabeth Robling Denning. *Cryptography and data security*. Addison-Wesley Reading, 1982.

[DHFK18] Mario Dobrovnik, David M Herold, Elmar Fürst, and Sebastian Kummer. Blockchain for and in Logistics: What to Adopt and Where to Start. *Logistics*, 2(3):18, 2018.

[Dir95] E. U. Directive. 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the EC*, 1995.

[Dir16] NIS Directive. Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union. *OJ L*, page 2016, 2016.

[dledL18] Commission Nationale de l'Informatique et des Libertés. Blockchain and the GDPR: Solutions for a responsible use of the blockchain in the context of personal data. `https://www.cnil.fr/en/blockchain-and-gdpr-solutions-responsible-use-blockchain-context-personal-data`, November 2018. (Accessed on 04/11/2020).

[DLZ$^+$18] Tien Tuan Anh Dinh, Rui Liu, Meihui Zhang, Gang Chen, Beng Chin Ooi, and Ji Wang. Untangling blockchain: A data processing view of blockchain systems. *IEEE transactions on knowledge and data engineering*, 30(7):1366–1385, 2018.

[DMMRP19] Beniamino Di Martino, Alfonso Marino, Massimiliano Rak, and Paolo Pariso. Optimization and Validation of eGovernment Business Processes with Support of Semantic Techniques. In *Conference on Complex, Intelligent, and Software Intensive Systems*, pages 827–836. Springer, 2019.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[DMS05] Roger Dingledine, Nick Mathewson, and Paul Syverson. Challenges in deploying low-latency anonymity. *NRL CHACS Report*, pages 5540–625, 2005.

[DN92] Cynthia Dwork and Moni Naor. Pricing via processing or combatting junk mail. In *Annual international cryptology conference*, pages 139–147. Springer, 1992.

[DP98] Thomas H. Davenport and Laurence Prusak. *Working Knowledge: How Organizations Manage what They Know*. Harvard Business School Press, 1998.

[Dro20] A. Drożdż. *Protection of Natural Persons with Regard to Automated Individual Decision-Making in the GDPR*. Wolters Kluwer, 2020.

[DTF19] Mehmet Demir, Ozgur Turetken, and Alexander Ferworn. Blockchain based transparent vehicle insurance management. In *2019 Sixth International Conference on Software Defined Systems (SDS)*, pages 213–220. IEEE, 2019.

[EAHL16] Ariel Ekblaw, Asaph Azaria, John D Halamka, and Andrew Lippman. A Case Study for Blockchain in Healthcare: MedRec prototype for electronic health records and medical research data. In *Proceedings of IEEE open & big data conference*, page 13, 2016.

[EDG09] Nathan S Evans, Roger Dingledine, and Christian Grothoff. A Practical Congestion Attack on Tor Using Long Paths. In *USENIX Security Symposium*, pages 33–50, 2009.

[Edw16] Lilian Edwards. Privacy, security and data protection in smart cities: A critical EU law perspective. *Eur. Data Prot. L. Rev.*, page 28, 2016.

[EERM15] Khaled El Emam, Sam Rodgers, and Bradley Malin. Anonymising and sharing individual patient data. *bmj*, page h1139, 2015.

[EHK+16] B Evers, J Hols, E Kula, J Schouten, M Den Toom, RM van der Laan, and JA Pouwelse. Thirteen Years of Tor Attacks, 2016.

[Ema13] K.E. Emam. *Guide to the De-Identification of Personal Health Information*. CRC Press, 2013.

[Eur15] European Data Protection Supervisor. Meeting the challenges of big data, 2015.

[Eur18] European Parliament and the Council of the European Union. Directive 2018/1673. `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L1673`, 2018.

[Eur19a] European Data Protection Supervisor. Introduction to the hash function as a personal data pseudonymisation technique, 2019.

[Eur19b] European Parliament and Council of European Union. Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union. `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:250:FIN`, 2019.

[Fas17] Joachim Fasching. Anwendungsbereiche und ausgewählte Rechtsfragen der Blockchain-Technologie. 2017.

[FCFL18] Tiago M Fernández-Caramés and Paula Fraga-Lamas. A Review on the Use of Blockchain for the Internet of Things. *IEEE Access*, pages 32979–33001, 2018.

[FERES+14] O.S. Faragallah, E.S.M. El-Rabaie, F.E.A. El-Samie, A.I. Sallam, and H.S. El-Sayed. *Multilevel Security for Relational Databases*. Taylor & Francis, 2014.

[Fin19] Michèle Finck. *Blockchain and the General Data Protection Regulation: Can Distributed Ledgers Be Squared with European Data Protection Law?* European Parliament, 2019.

[Fir78] D Firnberg. To be Effective, Privacy Needs Security! In *Current Topics in Cybernetics and Systems*, pages 13–15. Springer, 1978.

[FKP15] Michael Fleder, Michael S. Kester, and Sudeep Pillai. Bitcoin Transaction Graph Analysis. *arXiv preprint arXiv:1502.01657*, 2015.

[Fli95] Uwe Flick. *Handbuch qualitative Sozialforschung Grundlagen, Konzepte, Methoden und Anwendungen*. Beltz, 1995.

[Flo14] Luciano Floridi. Open data, data protection, and group privacy. *Philosophy & Technology*, pages 1–3, 2014.

[FP20] Michèle Finck and Frank Pallas. They who must not be identified—distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law*, 2020.

[FS19] David Florysiak and Alexander Schandlbauer. The information content of ico white papers. *Available at SSRN 3265007*, 2019.

[(GA06] United States Government Accountability Office (GAO). *Personal Information: Key Federal Privacy Laws Do Not Require the Information Resellers to Safeguard All Sensitive Data*. DIANE Publishing, 2006.

[GCKG14] Arthur Gervais, Srdjan Capkun, Ghassan O Karame, and Damian Gruber. On the privacy provisions of bloom filters in lightweight bitcoin clients. In *Proceedings of the 30th Annual Computer Security Applications Conference*, pages 326–335, 2014.

[Gen09] Craig Gentry. *A fully homomorphic encryption scheme*. Stanford university, 2009.

[Ger78] Robert S Gerstein. Intimacy and privacy. *Ethics*, pages 76–81, 1978.

[GI17] Jennifer M Gabany and Kamal MF Itani. Data Security. In *Clinical Trials Design in Operative and Non Operative Invasive Procedures*, pages 303–311. Springer, 2017.

[Glo14] Christina Glon. Data Protection in the European Union: A Closer Look at the Current Patchwork of Data Protection Laws and the Proposed Reform That Could Replace Them All. *International Journal of Legal Information*, pages 471–492, 2014.

[Gof63] Erving Goffman. Stigma: Notes on the Management of Spoiled Identity, Prentice Hall. *Englewood Cliffs, New Jersey*, 1963.

[GRS96] David M Goldschlag, Michael G Reed, and Paul F Syverson. Hiding routing information. In *International workshop on information hiding*, pages 137–150. Springer, 1996.

[GS17] B.G. Glaser and A.L. Strauss. *Discovery of Grounded Theory: Strategies for Qualitative Research*. Taylor & Francis, 2017.

[Ham16] Volker Hammer. DIN 66398. *Datenschutz und Datensicherheit-DuD*, 40(8):528–533, 2016.

[HHL94] Chung-Ming Huang, Jenq-Muh Hsu, and Huei-Yang Lai. A modified transition tour protocol test method. *Journal of Systems Integration*, pages 257–300, 1994.

[Hie09] Claudia Hienerth. *Wissenschaftliches Arbeiten kompakt Bachelor- und Masterarbeiten erfolgreich erstellen*. Linde, 2009.

[HLD+18] Yue Hao, Yi Li, Xinghua Dong, Li Fang, and Ping Chen. Performance analysis of consensus algorithm in private blockchain. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 280–285. IEEE, 2018.

[HM20] Dominik Huth and Florian Matthes. ProPerData-A process model to support GDPR compliance. 2020.

[Hol20] Erik Hollnagel. Managing for Security. In *International Security Management*, pages 43–53. Springer, 2020.

[HR17] Garrick Hileman and Michel Rauchs. Global cryptocurrency benchmarking study. *Cambridge Centre for Alternative Finance*, pages 33–113, 2017.

[HVR+20] Ivan Homoliak, Sarad Venugopalan, Daniël Reijsbergen, Qingze Hum, Richard Schumi, and Pawel Szalachowski. The Security Reference Architecture for Blockchains: Towards a Standardized Model for Studying Vulnerabilities, Threats, and Defenses. *IEEE Communications Surveys & Tutorials*, 2020.

[IEE90] IEEE. IEEE Standard Glossary of Software Engineering Terminology, 1990.

[Ins21] Project Management Institute. *Guide to the Project Management Body of Knowledge (PMBOK Guide) and the Standard for Project Management*. Pmbok Guide. Project Management Institute, 2021.

[IP18] Nabil El Ioini and Claus Pahl. A Review of Distributed Ledger Technologies. In *Lecture Notes in Computer Science*, pages 277–288. Springer International Publishing, 2018.

[Isl17] Michael Isler. Datenschutz auf der Blockchain. *Jusletter IT*, December 2017.

[ISO13a] ISO/IEC. ISO/IEC 27001:2013 Information technology — Security techniques — Information security management systems — Requirements, 2013.

[ISO13b] ISO/IEC. ISO/IEC 27005:2013 Information technology — Security techniques — Information security risk management, 2013.

[ISO18] ISO/IEC. ISO/IEC 27000:2018 Information technology — Security techniques — Information security management systems — Overview and vocabulary, 2018.

[ISS+13] S. Ingolfo, A. Siena, A. Susi, A. Perini, and J. Mylopoulos. Modeling laws with nomos 2. In *2013 6th International Workshop on Requirements Engineering and Law (RELAW)*, pages 69–71, 2013.

[(IT12] International Telecommunication Union (ITU). ITU-T Y.4000/Y.2060 (06/2012). *Overview of the Internet of things (06/2012)*, June 2012.

[Jev83] W.S. Jevons. *Money & the Mechanism of Exchange*. The Humboldt library of science, no. 50-51. Humboldt publishing Company, 1883.

[JKCP15] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. Likert scale: Explored and explained. *British journal of applied science & technology*, 7(4):396, 2015.

[Joh99] Norman L Johnson. Diversity in decentralized systems: Enabling self-organizing solutions. In *Decentralization II Conference, UCLA*. Citeseer, 1999.

[KAJS16] Rainer Knyrim, Eva Angerler, Andrea Jelinek, and Katharina Schmidt. *Datenschutz-Grundverordnung: Praxishandbuch*. MANZ, 2016.

[Ker16] Wolfgang Kerber. Digital markets, data, and privacy: competition law, consumer law and data protection. *Journal of Intellectual Property Law & Practice*, pages 856–866, 2016.

[KKP+19] Oleksandr Kurbatov, Pavel Kravchenko, Nikolay Poluyanenko, Oleksiy Shapoval, and Tetiana Kuznetsova. Using ring signatures for an anonymous e-voting system. In *2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT)*, pages 187–190. IEEE, 2019.

[KKZK12] Rafiullah Khan, Sarmad Khan, Rifaqat Zaheer, and Shahid Khan. Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges. In *10th International Conference on Frontiers of Information Technology, FIT 2012, Islamabad, Pakistan, December 17-19, 2012*, pages 257–260, 2012.

[KL14] Bert-Jaap Koops and Ronald Leenes. Privacy regulation cannot be hard-coded. A critical comment on the 'privacy by design' provision in data-protection law. *International Review of Law, Computers & Technology*, pages 159–171, 2014.

[KMH10] S. Keeney, H. McKenna, and F. Hasson. *The Delphi Technique in Nursing and Health Research*. Wiley, 2010.

[Koo19] Maria Koomen. The Encryption Debate in the European Union. *Carnegie Endowment for International Peace*, 2019.

[KS19] Kiran Kumar Kondru and Rajiakodi Saranya. Directed acyclic graph-based distributed ledgers—an evolutionary perspective. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(1), 2019.

[KTKT19] Evgenia Kapassa, Marios Touloupos, Dimosthenis Kyriazis, and Marinos Themistocleous. A smart distributed marketplace. In *European, Mediterranean, and Middle Eastern Conference on Information Systems*, pages 458–468. Springer, 2019.

[Kup20] Michael Kuperberg. Towards Enabling Deletion in Append-Only Blockchains to Support Data Growth Management and GDPR Compliance. In *2020 IEEE International Conference on Blockchain (Blockchain)*, pages 393–400. IEEE, 2020.

[Lam16] Siegfried Lamnek. *Qualitative Sozialforschung*. Beltz, 2016.

[Lau21] K. Lauter. *Protecting Privacy through Homomorphic Encryption*. 2021.

[LCNJ20] Joon Ho Lim, Ji Young Chun, Geontae Noh, and Ik Rae Jeong. GDPR Compliant Blockchain Based Access Control (GCBAC). *Journal of the Korea Institute of Information Security & Cryptology*, 30(6):981–997, 2020.

[Lew13] James A Lewis. *Raising the bar for cybersecurity*. Center for Strategic and International Studies, 2013.

[LGO20] Isabel Maria Lopes, Teresa Guarda, and Pedro Oliveira. General Data Protection Regulation in Health Clinics. *Journal of Medical Systems*, page 53, 2020.

[LLH$^+$20] Laphou Lao, Zecheng Li, Songlin Hou, Bin Xiao, Songtao Guo, and Yuanyuan Yang. A survey of IoT applications in blockchain systems: Architecture, consensus, and traffic modeling. *ACM Computing Surveys (CSUR)*, pages 1–32, 2020.

[LN20] Jerome Le Ny. *Differential Privacy for Dynamic Data*. SpringerBriefs in Electrical and Computer Engineering. Springer International Publishing, 2020.

[LRWW04] Brian N Levine, Michael K Reiter, Chenxi Wang, and Matthew Wright. Timing attacks in low-latency mix systems. In *International Conference on Financial Cryptography*, pages 251–265. Springer, 2004.

[LWH20] Aiya Li, Xianhua Wei, and Zhou He. Robust proof of stake: A new consensus protocol for sustainable blockchain systems. *Sustainability*, page 2824, 2020.

[LXCL17] Sin Kuang Lo, Xiwei Xu, Yin Kia Chiam, and Qinghua Lu. Evaluating suitability of applying blockchain. In *2017 22nd International Conference on Engineering of Complex Computer Systems (ICECCS)*, pages 158–161. IEEE, 2017.

[LYOK19] Nam-Yong Lee, Jinhong Yang, Md Mehedi Hassan Onik, and Chul-Soo Kim. Modifiable public blockchains using truncated hashing and sidechains. *IEEE Access*, 7:173571–173582, 2019.

[May09] Horst Mayer. *Interview und schriftliche Befragung Entwicklung, Durchführung und Auswertung*. Oldenbourg, 2009.

[May16] Philipp Mayring. *Einführung in die qualitative Sozialforschung eine Anleitung zu qualitativem Denken*. Beltz, 2016.

[McC18] Emma McClarkin. on Blockchain: a forward-looking trade policy (2018/2085(INI)). `https://www.europarl.europa.eu/doceo/document/A-8-2018-0407_EN.pdf`, November 2018. (Accessed on 04/01/2021).

[MGM+17] Roger Maull, Phil Godsiff, Catherine Mulligan, Alan Brown, and Beth Kewell. Distributed ledger technology: Applications and implications. *Strategic Change*, pages 481–489, 2017.

[Mic13] MG Michael. *Uberveillance and the Social Implications of Microchip Implants: Emerging Technologies: Emerging Technologies*. IGI Global, 2013.

[Mit09] Michael Mitzenmacher. Bloom Filters. In *Encyclopedia of Database Systems*, pages 252–255. Springer, 2009.

[MM02] Petar Maymounkov and David Mazieres. Kademlia: A peer-to-peer information system based on the xor metric. In *International Workshop on Peer-to-Peer Systems*, pages 53–65. Springer, 2002.

[MMLN17] Andrew Miller, Malte Möser, Kevin Lee, and Arvind Narayanan. An Empirical Analysis of Linkability in the Monero Blockchain. *arXiv preprint arXiv:1704.04299*, 2017.

[MOR01] Silvio Micali, Kazuo Ohta, and Leonid Reyzin. Accountable-subgroup multisignatures. In *Proceedings of the 8th ACM Conference on Computer and Communications Security*, pages 245–254, 2001.

[MPJ+13] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140, 2013.

[MS19] Lukas Müller and Reto Seiler. Smart contracts aus sicht des vertragsrechtsfunktionsweise, anwendungsfälle und leistungsstörungen. *Aktuelle Juristische Praxis*, 27(3):317–328, 2019.

[MSP15] Viktor Mayer-Schonberger and Yann Padova. Regime change: enabling big data through Europe's new data protection regulation. *Colum. Sci. & Tech. L. Rev.*, page 315, 2015.

[MVHA07] E.R. Monsen, L. Van Horn, and American Dietetic Association. *Research: Successful Approaches*. American Dietetic Association, 2007.

[MvOV18] A.J. Menezes, P.C. van Oorschot, and S.A. Vanstone. *Handbook of Applied Cryptography*. Discrete Mathematics and Its Applications. CRC Press, 2018.

[MW17] Mario Martini and Quirin Weinzierl. Die Blockchain-Technologie und das Recht auf Vergessenwerden: Zum Dilemma zwischen Nicht-Vergessen-Können und Vergessen-Müssen. *Neue Zeitschrift für Verwaltungsrecht*, pages 1251–1259, 2017.

[Nak08] Satoshi Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System*. Satoshi Nakamoto Institute, 2008.

[NHH07] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs*, pages 100–126, 2007.

[NM05] Graeme R Newman and Megan M McNally. Identity theft literature review. 2005.

[NS08] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE, 2008.

[Off18] Adrian Offerman. Swiss City of Zug issues Ethereum blockchain-based eIDs. `https://joinup.ec.europa.eu/collection/egovernment/document/swiss-city-zug-issues-ethereum-blockchain-based-eids`, February 2018. (Accessed on 03/27/2021).

[oHHS12] U.S. Department of Health & Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. 2012.

[Oku14] Krzysztof Okupski. Bitcoin developer reference. Technische Universiteit Eindhoven, 2014.

[Par10] Data Protection Working Party. Opinion 1/2010 on the concepts of "controller" and "processor". *Ref. WP*, 2010.

[Par14] Article 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques (WP 216). 2014.

[Par17a] Article 29 Data Protection Working Party. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679 . 2017.

[Par17b] Article 29 Data Protection Working Party. Guidelines on Data Protection Officers (DPOs). *Ref. WP*, 2017.

[Par17c] Article 29 Data Protection Working Party. Guidelines on the right to data portability, adopted on 13 December 2016. *Datenschutz und Datensicherheit-DuD*, pages 136–136, 2017.

[Pat20] LexisNexis PatentSight. Largest blockchain patent owners in Europe as of November 2020, by number of filings at the European Patent Office. Survey, Lexis Nexis, November 2020.

[PBSBS21] Katarina Preikschat, Moritz Böhmecke-Schwafert, Jan-Paul Buchwald, and Carolin Stickel. Trusted systems of records based on Blockchain technology-a prototype for mileage storing in the automotive industry. *Concurrency and Computation: Practice and Experience*, 33(1):e5630, 2021.

[Pec17] Morgen E Peck. Blockchain world-Do you need a blockchain? This chart will tell you if the technology can solve your problem. *IEEE Spectrum*, pages 38–60, 2017.

[Pei20] Carlo Peitz. *Datenschutzrechtliche Verantwortlichkeit in Blockchain-Systemen*. Springer, 2020.

[PKS21] Young-Hoon Park, Yejin Kim, and Junho Shim. Blockchain-based privacy-preserving system for genomic data management using local differential privacy. *Electronics*, 10(23):3019, 2021.

[PMR18] Guido Perboli, Stefano Musso, and Mariangela Rosano. Blockchain in logistics and supply chain: A lean approach for designing real-world use cases. *Ieee Access*, 6:62018–62028, 2018.

[PRB19] Asger B Pedersen, Marten Risius, and Roman Beck. A ten-step decision path to determine when to use blockchain technologies. *MIS Quarterly Executive*, pages 99–115, 2019.

[PST17] Suporn Pongnumkul, Chaiyaphum Siripanpornchana, and Suttipong Thajchayapong. Performance analysis of private blockchain platforms in varying workloads. In *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–6. IEEE, 2017.

[PTAA16] Danica Porobic, Pınar Tözün, Raja Appuswamy, and Anastasia Ailamaki. More than a network: distributed OLTP on clusters of hardware islands. In *Proceedings of the 12th International Workshop on Data Management on New Hardware*, pages 1–8, 2016.

[PTF16] Jules Polonetsky, Omer Tene, and Kelsey Finch. Shades of Gray: Seeing the Full Spectrum of Practical Data De-Intentification. *Santa Clara L. Rev.*, page 593, 2016.

[PTMT19] Pille Pullonen, Jake Tom, Raimundas Matulevičius, and Aivo Toots. Privacy-enhanced BPMN: Enabling data privacy analysis in business processes models. *Software and Systems Modeling*, 18(6):3235–3264, 2019.

[PTRC07] Ken Peffers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.

[PV19] Christian M Piska and Oliver Völkel, editors. *Blockchain rules.* Handbuch [Manz]. Manz Verlag, Wien, 2019.

[R$^+$09] Roxana Maria Roba et al. The Legal Protection Of The Secrecy Of Correspondence. *Curentul Juridic, The Juridical Current, Le Courant Juridique*, pages 135–154, 2009.

[Rau14] Judith Rauhofer. Round and Round the Garden? Big Data, Small Government and the Balance of Power in the Information Age. *University of Edinburgh, School of Law, Working Papers*, 2014.

[Rep19] Joint Report. First report of the observatory function on encryption. *European Union*, 2019.

[RH12] Fergal Reid and Martin Harrigan. An Analysis of Anonymity in the Bitcoin System. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 1318–1326, 2012.

[RH20] Frank Romeike and Peter Hager. Risiko-Management in der Informations- und Kommunikationstechnologie (IuK). In *Erfolgsfaktor Risiko-Management 4.0*, pages 479–520. Springer, 2020.

[RLWK21] Stephan Ramesohl, Julian Lauten-Weiss, and Georg Kobiela. Blockchains nachhaltig gestalten: Vorschlag von nachhaltigkeitsorientierten Entscheidungskriterien und eines Verfahrenskonzepts für die Umsetzung staatlich geförderter oder initiierter Projekte im Bereich Blockchain. Technical report, Wuppertal Report, 2021.

[Ros14] Agustín Rossi. Internet Privacy: Who Sets the Global Standard? *The International Spectator*, pages 65–80, 2014.

[RP09] Martin Rost and Andreas Pfitzmann. Datenschutz-schutzziele—revisited. *Datenschutz und Datensicherheit-DuD*, 33(6):353–358, 2009.

[RR14] Rachel V Rose and Lance H Rose. Appreciating healthcare data privacy laws in Canada, the United Kingdom, and the United States. *EDPACS*, pages 18–24, 2014.

[RS13] Dorit Ron and Adi Shamir. Quantitative Analysis of the Full Bitcoin Transaction Graph. In *International Conference on Financial Cryptography and Data Security*, pages 6–24, 2013.

[RSG98] Michael G Reed, Paul F Syverson, and David M Goldschlag. Anonymous connections and onion routing. *IEEE Journal on Selected areas in Communications*, 16(4):482–494, 1998.

[RST01] Ronald L Rivest, Adi Shamir, and Yael Tauman. How to leak a secret. In *International conference on the theory and application of cryptology and information security*, pages 552–565. Springer, 2001.

[Ruz20] Niederbacher Ruzicka. Deloitte Umfrage, Bestandsaufnahme nach 18 Monaten EU-DSGVO. Survey, Deloitte, 2020.

[RVC19] Tiago Rosado, André Vasconcelos, and Miguel Correia. A blockchain use case for car registration. In *Book Essentials Blockchain Technology*, pages 205–234. CRC Press, 2019.

[SA21] Ravital Solomon and Ghada Almashaqbeh. smartFHE: Privacy-Preserving Smart Contracts from Fully Homomorphic Encryption. *Cryptology ePrint Archive*, 2021.

[SABH15] Sarah Spiekermann, Alessandro Acquisti, Rainer Böhme, and Kai-Lung Hui. The challenges of personal data markets and privacy. *Electronic markets*, pages 161–167, 2015.

[Sal20] Fahad Saleh. Blockchain without waste: Proof-of-stake. *Available at SSRN 3183935*, 2020.

[SAWH14] Ali Seyed Shirkhorshidi, Saeed Aghabozorgi, Teh Ying Wah, and Tutut Herawan. Big data clustering: a review. In *International conference on computational science and its applications*, pages 707–720. Springer, 2014.

[SB17] Fabian Schär and Aleksander Berentsen. *Bitcoin, Blockchain und Kryptoassets: Eine umfassende Einführung.* Books on Demand, 2017.

[SBW17] Daniel Schatz, Rabih Bashroush, and Julie Wall. Towards a more representative definition of cyber security. *Journal of Digital Forensics, Security and Law*, pages 53–74, 2017.

[Sch10] Peter Schaar. Privacy by Design. *Identity in The Information Society*, pages 267–274, 2010.

[SDG14] Mattia Salnitri, Fabiano Dalpiaz, and Paolo Giorgini. Modeling and verifying security policies in business processes. In *Enterprise, business-process and information systems modeling*, pages 200–214. Springer, 2014.

[SFN$^+$18] D. Schmelz, G. Fischer, P. Niemeier, L. Zhu, and T. Grechenig. Towards Using Public Blockchain in Information-Centric Networks: Challenges Imposed by the European Union's General Data Protection Regulation. In *Proceedings of 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN 2018)*, pages 223–228, 2018.

[SGR99] Paul Syverson, D Goldschlag, and M Reed. Onion routing for anonymous and private internet connections. *Communications of the ACM*, 42(2):5, 1999.

[SHN$^+$18] Kongrath Suankaewmanee, Dinh Thai Hoang, Dusit Niyato, Suttinee Sawadsitang, Ping Wang, and Zhu Han. Performance analysis and application of mobile blockchain. In *2018 international conference on computing, networking and communications (ICNC)*, pages 642–646. IEEE, 2018.

[Sol08] Daniel J. Solove. *Understanding privacy.* Harvard University Press, 2008.

[SPB$^+$20] D. Schmelz, K. Pinter, J. Brottrager, P. Niemeier, R. Lamber, and T. Grechenig. Securing the rights of data subjects with blockchain technology. 2020.

[SPNG21] Dominik Schmelz, Karl Pinter, Phillip Niemeier, and Thomas Grechenig. Towards Informational Self-determination: Data Portability Requests Based on GDPR by Providing Public Platforms for Authorised Minimal Invasive Privacy Protection. In *International Congress on Blockchain and Applications*, pages 106–116. Springer, 2021.

[SPS$^+$19] Dominik Schmelz, Karl Pinter, Stefan Strobl, Lei Zhu, Phillip Niemeier, and Thomas Grechenig. Technical mechanics of a trans-border waste flow tracking solution based on blockchain technology. In *2019 IEEE 35th international conference on data engineering workshops (ICDEW)*, pages 31–36. IEEE, 2019.

[SS03] Andrei Serjantov and Peter Sewell. Passive attack analysis for connection-based anonymity systems. In *European Symposium on Research in Computer Security*, pages 116–131. Springer, 2003.

[SS14] Paul M Schwartz and Daniel J Solove. Reconciling personal information in the United States and European Union. *Calif. L. Rev.*, page 877, 2014.

[SS16] Gerald Spindler and Philipp Schmechel. Personal data and encryption in the European general data protection regulation. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, page 163, 2016.

[Sta19] Statista. The Big Business of Big Data - global big data and business analytics revenue between 2015 and 2022. `https://www.statista.com/chart/18328/big-data-business-analytics-revenue/`, June 2019.

[Ste10] A. Stevenson. *Oxford Dictionary of English*. Oxford Dictionary of English. OUP Oxford, 2010.

[Sti05] Hubert Stigler. *Praxisbuch empirische Sozialforschung in den Erziehungs- und Bildungswissenschaften*. Studien-Verlag, 2005.

[Str19] P. E. Strandberg. Ethical Interviews in Software Engineering. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 1–11, 2019.

[Swe00] Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, pages 1–34, 2000.

[Syd18] Gernot Sydow. *Europäische Datenschutzgrundverordnung*. Nomos, 2018.

[Tak17] Koji Takahashi. Blockchain Technology for Letters of Credits and Escrow Arrangements. 2017.

[Tar10] S. Tarkoma. *Overlay Networks: Toward Information Networking*. CRC Press, 2010.

[Tik18] Sergei Tikhomirov. *Ethereum: State of Knowledge and Research Perspectives*, pages 206–221. 2018.

[TPRM18] Christina Tikkinen-Piri, Anna Rohunen, and Jouni Markkula. EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, pages 134–153, 2018.

[TS16] Florian Tschorsch and Bjorn Scheuermann. Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies. *IEEE Communications Surveys and Tutorials*, pages 2084–2123, 2016.

[TTM+15] Sabine Trepte, Doris Teutsch, Philipp K Masur, Carolin Eicher, Mona Fischer, Alisa Hennhöfer, and Fabienne Lind. Do people know about privacy and data protection strategies? Towards the "Online Privacy Literacy Scale". In *Reforming European data protection law*, pages 333–365. Springer, 2015.

[TvS18] A.S. Tanenbaum and M. van Steen. *Distributed Systems - Third edition.* Pearson Education, Inc., 2018.

[UNC16] Data Protection Regulation UNCTAD. International data flows: Implications for trade and development. *United Nations*, 2016.

[VBBO03] GW Van Blarkom, John J Borking, and JG Eddy Olk. Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague*, 2003.

[VMH06] G.F. Vito, J.R. Maahs, and R.M. Holmes. *Criminology: Theory, Research, and Policy.* Criminal justice illuminated. Jones and Bartlett, 2006.

[VVdB17] Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 2017.

[W+16] MGCSA Walport et al. Distributed ledger technology: Beyond blockchain. *UK Government Office for Science*, pages 1–88, 2016.

[WA16] Martin A Weiss and Kristin Archick. US-EU data privacy: from safe harbor to privacy shield, 2016.

[Wai17] Alexa Wainscott. A Golden Key to Pandora's Box: The Security Risks of Government-Mandated Backdoors to Encrypted Communications. *N. Ky. L. Rev.*, page 57, 2017.

[Wal18] Ari Ezra Waldman. *Privacy as trust: Information privacy for an information age.* Cambridge University Press, 2018.

[Wan11] H. Wang. *Protecting Privacy in China: A Research on China's Privacy Standards and the Possibility of Establishing the Right to Privacy and the Information Privacy Protection Legislation in Modern China.* Springer Berlin Heidelberg, 2011.

[WFMC17] K. Williams, J.M. Facciola, P. McCann, and V.M. Catanzaro. *The Legal Technology Guidebook.* Springer International Publishing, 2017.

[Wol78] Maxine Wolfe. Childhood and privacy. In *Children and the environment*, pages 175–222. Springer, 1978.

[Won07] Rebecca Wong. Data protection online: alternative approaches to sensitive data. *J. Int'l Com. L. & Tech.*, 2007.

[WSL+19] Licheng Wang, Xiaoying Shen, Jing Li, Jun Shao, and Yixian Yang. Cryptographic primitives in blockchains. *Journal of Network and Computer Applications*, 127:43–58, 2019.

[WWC$^+$18] Hai Wang, Yong Wang, Zigang Cao, Zhen Li, and Gang Xiong. An overview of blockchain security analysis. In *China Cyber Security Annual Conference*, pages 55–72. Springer, Singapore, 2018.

[XWS19] Xiwei Xu, Ingo Weber, and Mark Staples. *Architecture for blockchain applications*. Springer, 2019.

[Yak18] Anatoly Yakovenko. Solana: A new architecture for a high performance blockchain v0. 8.13. *Whitepaper*, 2018.

[ZCW$^+$14] Zhi-Kai Zhang, Michael Cheng Yi Cho, Chia-Wei Wang, Chia-Wei Hsu, Chong-Kuan Chen, and Shiuhpyng Shieh. IoT security: ongoing challenges and research opportunities. In *2014 IEEE 7th international conference on service-oriented computing and applications*, pages 230–234. IEEE, 2014.

[ZFG$^+$04] Ye Zhu, Xinwen Fu, Bryan Graham, Riccardo Bettati, and Wei Zhao. On flow correlation attacks and countermeasures in mix networks. In *International Workshop on Privacy Enhancing Technologies*, pages 207–225. Springer, 2004.

[ZMVOM18] Xiaochen Zheng, Raghava Rao Mukkamala, Ravi Vatrapu, and Joaqun Ordieres-Mere. Blockchain-based personal health data sharing system using cloud storage. In *2018 IEEE 20th international conference on e-health networking, applications and services (Healthcom)*, pages 1–6. IEEE, 2018.

[ZZPZ19] Qingchuan Zhao, Chaoshun Zuo, Giancarlo Pellegrino, and Li Zhiqiang. Geo-locating Drivers: A Study of Sensitive Data Leakage in Ride-Hailing Services. In *Annual Network and Distributed System Security symposium*, 2019.

Web Ressources

[Beh21] Michael Behr. Facebook to access WhatsApp user data – except in Europe. https://digit.fyi/facebook-to-access-whatsapp-user-data-except-in-europe, January 2021. (Accessed on 2021-03-04).

[Bit20] Bitcoin Wiki. Protocol documentation. https://en.bitcoin.it/wiki/Protocol_documentation, December 2020. (Accessed on 2021-03-04).

[Blo21] Blockchain.com. Blockchain Charts - Average Block Size (MB). https://www.blockchain.com/charts/avg-block-size, April 2021. (Accessed on 2021-04-05).

[But17] Vitalik Buterin. Proof of stake FAQ. https://vitalik.ca/general/2017/12/31/pos_faq.html#how-does-validator-selection-work-and-what-is-stake-grinding, December 2017. (Accessed on 2021-12-29).

[Com18] European Commission. Anti-money laundering and countering the financing of terrorism. https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-supervision-and-risk-management/anti-money-laundering-and-countering-financing-terrorism_en, July 2018. (Accessed on 2021-12-09).

[Cza17] Jacek Czarnecki. Blockchains and Personal Data Protection Regulations Explained. https://www.coindesk.com/blockchains-personal-data-protection-regulations-explained, April 2017. (Accessed on 2020-10-09).

[ELOP20] Josh Elbahrawy, James Lovejoy, Anne Ouyang, and Justin Perez. Analysis of Bitcoin Improvement Proposal 340 — Schnorr Signatures. https://courses.csail.mit.edu/6.857/2020/projects/4-Elbahrawy-Lovejoy-Ouyang-Perez.pdf, May 2020. (Accessed on 2021-02-03).

[Fen18] Briana Feng. Decoding zk-SNARKs. https://medium.com/wolverineblockchain/decoding-zk-snarks-85e73886a040, February 2018. (Accessed on 2021-04-10).

[Fou21] Ethereum Foundation. Proof-of-stake (PoS). https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/, December 2021. (Accessed on 2021-12-29).

[Fox21] Matthew Fox. Businessinsider: Ethereum founder Vitalik Buterin breaks down why the London hard fork is so important for the future of ether. https://markets.businessinsider.com/news/currencies/ethereum-london-hard-fork-eip-1559-vitalik-buterin-carbon-emissions-2021-8, August 2021. (Accessed on 2021-12-11).

[Goo21] Google. Google Trends - Statistic of Worldwide Searches for Distributed Ledger. https://trends.google.com/trends/explore?date=2004-01-01%202022-02-17&q=%2Fg%2F11cn6flkr3, January 2021. (Accessed on 2021-05-18).

[Gro21] Satis Group. Cryptoasset market coverage initiation: Network creation. https://research.bloomberg.com/pub/res/d28giW28tf6G7T_Wr77aU0gDgFQ, July 2021. (Accessed on 2021-08-12).

[inc20] incoPat. Total number of blockchain patent applications filed globally as of April 2020, by leading company. https://www.statista.com/statistics/1119248/global-number-of-blockchain-patent-applications-filed-by-leading-company/, April 2020. (Accessed on 2021-02-24).

[Lop22] Jameson Lopp. Bitcoin Core Config Generator. https://jlopp.github.io/bitcoin-core-config-generator/, May 2022. (Accessed on 2022-03-02).

[ME21] David McCandless and Tom Evans. World's Biggest Data Breaches & Hacks — Information is Beautiful. https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/, April 2021. (Accessed on 2021-04-18).

[ndledl20] CNIL Commission nationale de l'informatique et des libertés. Sheet number 13: Prepare for the exercise of people's rights. https://www.cnil.fr/en/sheet-ndeg13-prepare-exercise-peoples-rights, June 2020. (Accessed on 2021-03-05).

[OMG14] BPMN OMG. Business Process Model and Notation. https://www.omg.org/spec/BPMN, January 2014. (Accessed on 2021-06-18).

[Ovi21] Shira Ovide. The Truth About Your WhatsApp Data. https://www.nytimes.com/2021/01/13/technology/whatsapp-data.html, January 2021. (Accessed on 2021-03-04).

[Pas21] Marta Pastor. EBSI Documentation. https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=381517902, July 2021. (Accessed on 2022-01-10).

[Pro] The Monero Project. Bulletproofs. https://web.getmonero.org/resources/moneropedia/bulletproofs.html. (Accessed on 2022-01-29).

[PRVB13] Marek Palatinus, Pavol Rusnak, Aaron Voisine, and Sean Bowe. Mnemonic code for generating deterministic keys. https://github.com/bitcoin/bips/blob/master/bip-0039.mediawiki, September 2013. (Accessed on 2021-04-11).

[QG22] Marco Quiroz-Gutierrez. Crypto is fully banned in china and 8 other countries. https://fortune.com/2022/01/04/crypto-banned-china-other-countries, January 2022. (Accessed on 2022-01-27).

[Rei16] Christian Reitwiessner. zkSNARKs in a nutshell. https://blog.ethereum.org/2016/12/05/zksnarks-in-a-nutshell/, December 2016. (Accessed on 2021-04-10).

[Res20] Finextra Research. Chinese giants lead blockchain patent applications 2020. https://www.finextra.com/newsarticle/35692/chinese-giants-lead-blockchain-patent-applications, April 2020. (Accessed on 2020-11-10).

[Sal20] Wigan Salazar. Statement from buchbinder. https://www.buchbinder.de/fileadmin/user_upload/corporate/user_upload/Pressemeldungen/2020_01_25_STATEMENT_FROM_BUCHBINDER_EN_final.pdf, 2020. (Accessed on 2021-07-02).

[Sie] Jerrold Siegel. Hashing Functions and Their Uses In Cryptography. http://www.umsl.edu/~siegelj/information_theory/projects/HashingFunctionsInCryptography.html. (Accessed on 2022-01-28).

[Smi13] Gerry Smith. Meet Tor, The Military-Made Privacy Network That Counts Edward Snowden As A Fan. https://www.huffpost.com/entry/tor-snowden_n_3610370, August 2013. (Accessed on 2021-01-03).

[Ste14] Joseph Steinberg. These Devices May Be Spying On You (Even In Your Own Home). https://www.forbes.com/sites/josephsteinberg/2014/01/27/these-devices-may-be-spying-on-you-even-in-your-own-home, January 2014. (Accessed on 2021-02-04).

[Sur21] SurveyMonkey. SurveyMonkey and GDPR - How we're helping our customers stay compliant. https://prod.smassets.net/assets/cms/sm/uploads//SurveyMonkey-GDPR-Whitepaper-Jul20.pdf, April 2021. (Accessed on 2022-03-05).

[Tod16] Peter Todd. OpenTimestamps: Scalable, Trustless, Distributed Timestamping with Bitcoin. https://petertodd.org/2016/opentimestamps-announcement, 2016. (Accessed on 2021-03-04).

[Tun22] Abi Tyas Tunggal. The 65 biggest data breaches. https://www.upguard.com/blog/biggest-data-breaches, 2022. (Accessed on 2022-02-02).

[Uni21] Tari Labs University. Bulletproofs and Mimblewimble. https://tlu.tarilabs.com/cryptography/bulletproofs-and-mimblewimble/MainReport.html, January 2021. (Accessed on 2021-04-10).

[VFK20] Daniel Victor, Sheera Frenkel, and Isabel Kershner. Personal Data of All 6.5 Million Israeli Voters Is Exposed. https://www.nytimes.com/2020/02/10/world/middleeast/israeli-voters-leak.html, February 2020. (Accessed on 2021-02-01).

[VS13] Nicolas Van Saberhagen. Cryptonote v 2.0. https://web.archive.org/web/2020*/https://cryptonote.org/whitepaper.pdf, 2013. (Archived, Accessed on 2021-02-02).

[Wui12] Pieter Wuille. Hierarchical Deterministic Wallets. https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki, February 2012. (Accessed on 2021-11-11).