T. van Biemen

Personal Privacy in Practice

Putting the GDPR to test in a collective exercise of data subjects' right of access.

TU Delft

Personal privacy in practice

Putting the GDPR to test in a collective exercise of strengthened data subjects access rights.

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering & Policy Analysis**

Faculty of Technology, Policy and Management

By

Thomas van Biemen

Student number: 4206827

to be defended publicly on Friday October 19, 2018 at 14:00.

Graduation committee:

Supervisor:	Dr. H. Asghari
Thesis committee:	Dr. M.E. Warnier

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

“The good news about privacy is that eighty-four percent of us are concerned about privacy. The bad news is that we do not know what we mean.” (Branscomb & Larson, 1995)

Preface

Beste lezer, Utrecht, 12 oktober 2018

Voor u vindt u de eindrapportage van het project waar ik de afgelopen zes maanden met veel plezier een enorme hoeveelheid tijd in heb geïnvesteerd. Toen ik zes maanden geleden door Dr. Asghari werd gevraagd om mijn afstudeerproject te richten op het raakvlak tussen privacy en ICT, was de klik met het onderwerp voor mij meteen duidelijk. Ik denk dat iedereen die dagelijks ervaart hoe soepel de analyse van een grote hoeveelheid gegevens is geïntegreerd in het moderne leven, in ieder geval onderbewust begrijpt dat dit raakvlak de nodige uitdagingen met zich meebrengt. De juridische wereld van het privacyrecht en de technische wereld van de data-analyse zijn in het oplossen van problemen zo verschillend dat een goede combinatie van de twee mij vanaf het begin van het project al bijna onmogelijk in de oren klonk. Binnen een week herschreef ik het onderzoeksvoorstel waar ik al een aantal maanden aan had gewerkt om tot een praktijkonderzoek naar de uitwerking van de AVG. Het uitvoeren van dit onderzoek bleek echter nog een stuk uitdagender dan ik op dat moment kon overzien.

De AVG kan ik het best uitleggen als een gereedschapskist die tot de rand gevuld is met complexe instrumenten die dienen om het fundamentele recht op privacy dat ieder mens verdient te verankeren in internationaal recht. Mijn onderzoek heeft zich gericht op de werking van slechts een van deze instrumenten in de praktijk. Toch denk ik met dit rapport wel degelijk mijn steentje te hebben bijgedragen aan het collectieve begrip van een belangrijk onderdeel van deze AVG. Ondanks alle commotie die is ontstaan rond de inwerktreding van de AVG, is er tot voor kort namelijk opvallend weinig informatie te vinden over het effect dat de wetgeving heeft op het dagelijks leven van Europese burgers wiens privacy het zou beschermen. Het mag duidelijk zijn dat dit effect ruimschoots gevonden is in dit onderzoek naar de manier waarop organisaties reageren op aanvragen tot inzicht van persoonlijke gegevens.

Ik ben enorm trots dat het is gelukt om binnen het geplande half jaar een rapport te kunnen produceren waarin duidelijke en goed onderbouwde bevindingen van deze uitwerking worden gepresenteerd. Toch merk ik dat het schrijven van deze laatste paragrafen als meer voelt dan het afronden van een project waar ik een half jaar hard aan heb gewerkt. In de analyse die ik heb kunnen uitvoeren op een complex probleem, het ontwerpen van een nieuwe dataverzamelingsmethode en de formulering van beleidsadvies die het uitvoeren van AVG-inzageverzoeken voor zowel organisaties als burgers zal verbeteren, merk ik ook de vaardigheden die ik de afgelopen zeven jaar in Delft heb geleerd. Voor u vindt u de eindrapportage van een project waarmee ik mijn bachelor en master Technische Bestuurskunde aan de Technische Universiteit Delft met trots kan afsluiten.

Wat dit project voor mij extra mooie afsluiting maakt, is de enorme hoeveelheid hulp die ik tijdens het proces heb ontvangen van iedereen die ik de afgelopen jaren heb leren kennen. Dit werd vooral duidelijk bij het zoeken van vrijwilligers namens wie ik inzage tot gegevens bij een grote hoeveelheid bedrijven kon aanvragen. Vrienden, vriendinnen en familieleden stonden allemaal klaar om inzicht in het gebruik van hun persoonlijke gegevens met mij te delen. Deze mensen wil ik hierbij nogmaals enorm bedanken voor hun hulp.

Bij het bedanken van iedereen die mij de afgelopen tijd heeft geholpen is het onmogelijk om de mensen over te slaan die mij vanaf het begin hebben begeleid. Dr. Asghari en Warnier hebben mij constant waardevol advies gegeven over de planning, afbakening, uitvoering en rapportage van mijn afstuderen. Het enige dat bij nader inzien op te merken is over dit advies is dat ik het allemaal beter ter harte had kunnen en soms moeten nemen. Dit betekent niet dat ik spijt heb over de keuzes die ik in de afgelopen zes maanden heb gemaakt. Ik ben er juist ook trots op dat ik deze keuzes steeds zelf gemaakt heb en daardoor met kan zeggen dat ik met mijn studietijd afsluit met een eigen eindrapportage van het meest uitdagende project dat ik in mijn leven tot nu toe heb uitgevoerd.

Thomas van Biemen

Executive summary

After a two-year grace period, the General Data Protection Regulation (GDPR) became enforceable on the 25th of May 2018. It presents the greatest update on European privacy regulation in over 20 years and aims to secure a fundamental human right in a modern Europe. This right to privacy is a public value that has proved hard to protect in the 21st century. It is no secret that the European Union has come a long way in creating this important piece of legislation. The GDPR affirms European citizen's right to informational self-determination. This grants individuals the authority to decide for themselves when and within what limits information about their private life should be shared with others (Rouvroy & Poullet, 2009) and protects individuals against "unlimited collection, storage, use and disclosure of his/her personal data" (Westin, 1970). In order to make this decision and potentially practice further personal privacy rights, Subject's Access Request (SARs) acts as a natural precursor to exercising the other rights by allowing subjects to gain insight into the processing that is performed using their personal data.

SARS have been a cornerstone of European privacy regulation for almost forty decades. However, research into exercising this right in practice has been sparse for many of those years. More recently, research by Norris et al. (2017) and others has shown a great disparity between the European data privacy law in writing and the application of the law in practice. Most academic literature written on the new GDPR is focused on the effect that certain passages (not) included in the final draft might have on practices within organizations. It is unclear however, if the regulation has increased the protection of fundamental rights, or if it just places an unnecessary burden on organizations. The introduction of the GDPR also provides a clear opportunity to understand the relationship that other factors, such as an organization's sector and size have on its response to SARs. This research aims to fill this knowledge gap and provide the first scientific insight into factors that influence the responses of organizations to Subject Access Requests under the GDPR, including the effect of this regulation itself.

Out of a sample of 116 Dutch organizations, 51% responded to SARs in a manner compliant with the GDPR. Although this means that non-compliance is still widespread, this share of compliant organizations is significantly higher than any other research performed on organizations under previous regulation. This difference is best illustrated in Table 1. The compliance rate that is shown in this table is based on organizations that both responded with a copy of the processed personal data and provided the required insight into the processing of this data by adequately answering further questions which are prescribed in the concerned regulation.

Study	N	Country	Sectors	Response rate	Response with data	Provided further insight	Compliance
Norris et al., 2017	183	EU	Mixed	80%	57%	43%	34%
Asghari, Mahieu, & van Eeten, 2017	106	The Netherlands	Mixed	83%	69%	27%	22%
Herrmann & Lindemann, 2016	120	Germany	Popular apps & websites	68%	n.a.	n.a.	43%
Spiller, 2016	17	UK	CCTV	n.a.	35%	n.a.	35%
Ausloos & Dewitte, 2018	60	EU	ISSS	74%	67%	n.a.	33%
GDPR part of van Biemen, 2018	116	The Netherlands	Mixed	81%	68%	65%	51%

Table 1. *Observed GDPR SAR compliance in this research, compared to compliance rates found in earlier research. Please see section 2.2.2 and 5.1 for a more detailed explanation and remarks concerning the comparison between these findings.*

Regression analysis on a combination of GDPR data collected for this research and data collected under the earlier Dutch wbp data privacy regulation shows an even bigger influence of the GDPR. When other factors are accounted for, the introduction of the GDPR has increased compliance rate by between 64% and 500%. Because of a small difference between the data collection performed for this research and the collection regime by Asghari et al. (2017), the real effect might even be even higher in reality. Besides the GDPR, the only factor which was shown to have a clear influence on SAR response compliance was the size of an organization. Contrary to expectations, smaller organizations clearly produce better responses to access requests, with a factor that is comparative to the effect that the GDPR has on compliance. This is probably because smaller organizations are more prone to invest time into a personal answer, a finding that is also supported by further qualitative analysis. Although differences in reaction also seem apparent between organizations operating in different sectors, no clear difference in compliance is found. Numerical analysis has shown a difference in the specificity of organizations' responses operating in the ICT and other services sectors. These responded in a noticeably more generic way to SARs compared to other organizations. Further qualitative analysis shows differences in responses from organizations in two other sectors. Governmental organizations show a distaste for electronic communication methods which would ease the process of requesting personal data. Organizations in the Dutch healthcare sector follow stricter and sometimes conflicting sector specific regulation which limits data subjects' rights to access certain information. The SAR response effect of other potentially relevant factors, such as organizations' operating locations or relationship with the data subject could not be established with the available data.

For many organizations, access requests in the GDPR era seem to no longer be a single incident, but a real possibility. The introduction of the GDPR has thus removed the barrier that was previously most often holding back the use of SARs in practice: it's obscurity. This difference in attitude can probably be attributed to the data protection hype that surrounded the introduction of the GDPR. And although this hype has also seemed to improve the attitude with which organizations process such requests, the data protection hype also has adverse effects on organizations' responses. These adverse effects are most apparent in two barriers that were often experienced during this research. The first is found in processes which are designed by predominantly larger organizations to deal with requests in a more streamlined way. These processes can confine data subjects in their requests which often only lead to only receiving partial insight into the processing of their personal data. The second barrier is found in excessive identification methods that are imposed by data controllers. Controllers often place large burdens on subjects to proof their identity, thereby requiring the processing of even more personal data. Organizations often seem anxious to protect the personal data in their possession but fail to see how their practices infringe on other privacy and GDPR fundamentals such as transparency and data minimalization.

Although the GDPR has definitely had a positive impact on the way organizations respond to access requests, the regulation does still not provide all subjects with information on the processing of personal data that is rightfully theirs. GDPR compliance in the field of access requests can be increased even further by facilitating an easier (digital) proof of identification and providing more detailed information for organizations on how (not) to respond to access requests. The current lack of these two, and the uncertainty that exists because of unclarified ambiguity in the regulation only feeds the paranoia that exists in some organizations with regards to the GDPR. Most non-compliant organizations seem to excessively protect information regarding the processing of personal data with the best intentions.

On the long term, it is important that the Data Protection Authority (DPA) shows that violations of the GDPR will indeed be punished. The anxiety that follows from these actions should be balanced with anxiety following from misunderstandings and ambiguities in the law. For this anxiety balance, it is vital that uncertainty over parts of the GDPR are constantly addressed. Especially when ambiguity is not yet addressed in court, the DPA should publish their detailed view on the matter. When punishing violators, the DPA should also detail what the penalized party should have done better, so that others can learn from their mistakes. In similar fashion, organizations should be assisted in optimizing their GDPR experience by the publication of best practices of organizations shown to be fully compliant. To achieve broader GDPR compliance, the Dutch government can also set a better example herself. Organizations in the public sector possess all tools required to design an optimal SAR experience, using electronic communication to disclose the processing of personal data to its fullest extent and using digital identifiers such as the DigID. Instead, most local governments seem to favor paper communication, restricting processes and obligatory identity checks in person.

By providing insight into the effect that the GDPR and organizational characteristics have on SAR responses, this research has set an important step in filling the knowledge gap that surrounds the exercise of GDPR access requests in practice. This thesis report features a detailed explanation of the literature background and data analysis that was performed to obtain these insights. The barriers and discourses that data subjects encountered during the execution of their newfound rights are described in detail and supplemented with policy advice in later chapters. Another important contribution that is included in this writing is the novel data collection method which uses the concept of distributed data collection to collect a bigger and more focused SAR response dataset with the help of volunteers. This research has set an important and sometimes challenging first step in using this method in practice. The decision of the Dutch DPO to specifically include this method as a possibility on their guide to exercising personal privacy rights is expected to increase both the awareness and convenience of collectively exercising personal privacy rights. The author is proud to present the highlighted scientific and societal contributions in more detail in the following chapters.

Contents

1 Introduction

The introduction of increasingly sophisticated data solutions has provided businesses and governments with immense opportunities. Every year, cheaper computers can store more data and process it using increasingly powerful algorithms. These advancements have not only resulted in incredibly successful new business models, powerful forecasting technology and huge new organizations but also changed the way people think about information (Bharadwaj et al., 2013; Davison, 2007). Only a few decades ago, data storage was expensive and would only be used to store necessary information. Statistic methods were developed to provide conclusive answers with the fewest data points possible. Nowadays, big data paradigms have organizations and governments collecting as much data as they can, believing that enough data will eventually lead to new extraordinary uses (Kitchin, 2014; Pokharel, 2013).

Big data and its predictive potential have indeed delivered promising results with applications in fields ranging from scientific research to national security and business exercises. (Krigsman, 2015; Marr, 2015; Naik & Joshi, 2017). However, the accompanied drive for growing data collection and application also brings its issues. (The Economist, 2017) Continuous data accumulation often means that organizations own more personal data than users assume, or organizations themselves even realize (Harford, 2014). In their quest for (processing) even greater data collections, both organizations and governments struggle with the protection and ethical use of their collected data, as exemplified by the Snowden revelations and more recently by the Cambridge Analytica scandal (Boyd & Crawford, 2012; Franceschi-Bicchierai, 2014; Greenfield, 2018; Yu & Cude, 2009).

1.1 Fundamental rights in a digital world

One of the biggest issues with the widespread use of big data and other ICT tools is the seeming disregard that some organizations have for their users' fundamental right to privacy (Hoepman, 2009; H. Jones & Soltren, 2005). Contrary to the US legal tradition of defining privacy as "the right to be left alone", privacy in the European Union in this context is supplemented by the right of informational self-determination. This right grants European individuals the authority to decide for themselves when and within what limits information about their private life should be shared with others (Rouvroy & Poullet, 2009) and protects individuals against "unlimited collection, storage, use and disclosure of his/her personal data" (Westin, 1970).

Safeguarding citizen's privacy rights has long been key to protecting important personal and democratic values of dignity, trust, self-development, autonomy and dissent. (Rouvroy & Poullet, 2009; Sunstein, 2003). Privacy, however, is also a right that is easily abandoned by the public in return for short term gains in other values such as safety or convenience. (Mansfield-Devine, 2015; SAS, 2015). Widespread examples of data protection failures and the popularity of data driven business models leads some to believe that "privacy is dead". (BBC News, 2017; Morgan, 2014; Sturges, 2005) However, although citizens' actions might not expose this, the right of privacy is still important to many, with consistently large numbers of consumers expressing privacy concerns in data collection practices (Ianelli, 2018; Gigya, 2017). Even when concerned, consumers seem unaware of both the amount of data that is being collected on them by organizations and government and the implications that this brings. Both organizations and governments appear to intentionally feed the information asymmetry that creates this unawareness by burying details of their data use in long privacy statements or vague regulations. (Martijn & Tokmetzis, 2016; Phelps, Nowak, & Ferrell, 2000)

Guaranteeing privacy and informational self-determination rights on a societal level is further restricted because regulation concerning the collection and processing of (personal) data often stem from times before the revolution of big data. Furthermore, these laws are often different between countries, while modern organizations operate in an increasingly worldwide environment. (Esteve, 2017) This made European privacy

laws ineffective in enforcing fundamental human rights, being either too lenient on new technology or unenforceable in their strictness. (Birnhack & Elkin-Koren, 2010).

The fundamental right to personal privacy, and balancing it's values with business and national security interests is identified as a key part of the grand challenge of achieving a secure cyberspace by the National Academy of Engineering (National Academy of Engineering, n.d.) and the United Nations development goal of peace, justice and strong institution (United Nations, 2017). Following these formulations, and the wickedness of the problem at hand (Alaqra, 2018; Levin, Cashore, Bernstein, & Auld, 2012) it is concluded that dealing with the problem at hand constitutes to the governance of international grand challenges (Borrás & Edler, 2015; Cagnin, Amanatidou, & Keenan, 2012; European Commission, 2016). Through analyzing and modelling the system surrounding this wicked problem, this thesis will follow the lessons learned in the master courses of the EPA program to contribute to both the scientific literature and the broader societal aspects of the problem. Conclusions are supplemented by recommendations for policy makers to govern the personal privacy aspect of the international grand challenge of privacy in a digital world.

1.2 Privacy governance

Discussions on salvaging privacy governance may have led to a dramatic change in privacy protection within the European Union. After years of negotiation and a further two-year grace period, the EU's General Data Protection Regulation (GDPR) has become enforceable in late May 2018. This regulation aims to strengthen data protection for all individuals within the European Economic Area by reforming and unifying regulation across the continent (Regulation (EU) 2016/679, 2016). The key idea that underpins the regulation is that data subjects still have rights concerning data that is collected on them, even though it is stored and processed by another party. (European Commission, 2018; F-Secure, 2018). This renewed fixation on informational self-determination is guarded by strengthened privacy rights, the right of Access, Rectification, Cancellation and Opposition to personal data processing. From these ARCO-rights, the right of access is the foremost important: subjects need to be able to know what data has been stored on them and how it is used before being able to make well informed decision on opposition or for rectification (Norris, de Hert, L'Hoiry, & Galetta, 2017).

The right of access is not a new tool. It has been included into EU member states law for quite some time, following EU directive 95/45/EC and in earlier form the 1981 convention 108. Although ARCO-rights have always theoretically guaranteed user's information self-determination, it has not yet delivered in practice. Studies that investigated its use under this previous regulation have described it as a failing instrument that is unknown by organizations, consumers and even judges, with widespread non-compliance often proceeding unpunished. (Asghari et al., 2017; Ausloos & Dewitte, 2018; Norris et al., 2017; Winter et al., 2008).

These results present a sobering outlook on the future of European digital privacy. However, the GDPR does have the potential to improve citizens fundamental rights to privacy as it is believed to provide stricter regulation when compared to its predecessors by proponents and opponents alike. (Albrecht, 2016; Allen, Berg, Berg, Markey-Towler, & Potts, 2018; U.S. Chamber, 2015). This potential is found not only in (marginal) changes to citizens' ARCO-rights, but also in the unification of regulation, increased power for privacy watchdogs and a general sense of urgency that the hype accompanying this new regulation brings to organizations. (Ausloos & Dewitte, 2018; Norris et al., 2017) Scholars disagree on the predicted strength of these effects and the consequences that it's application will have on ordinary citizens' ability to enforce their fundamental rights (Ausloos & Dewitte, 2018; Koops, 2014). Other effects, such as different responses by organizations operating in different sectors are also observed, but the extend of these effects has not yet been investigated. The goal of this thesis paper is to fill this knowledge gap by investigating how organizations and governments react to citizens exercising their privacy rights in practice under this new regulation, answering the following research question:

How are organizations responding to subject access requests under the GDPR, and what factors influence their responses?

To provide a conclusive and timely answer to this question, research will follow an alternative method compared to earlier investigations into data access requests in practice. Data obtained from requests made personally by researchers will be supplemented by answers received through requests made behalf of a group of 35 Dutch volunteers, to allow for greater choice in investigation differences in responses by organizations.

1.3 Outline

Further background to the research that was introduced in previous paragraphs can be found in the following chapter. The literature review that is presented in this chapter will start with detailing the history of privacy in a European context, before focusing on relevant digital privacy law and changes that the GDPR brings. In further sections, the chapter will discuss research into the exercise of data rights in practice and effects of the GDPR so far. The chapter will finish with the formulation of the main research question and its subquestions.

A third chapter will outline the methodology that will be used to answer these research questions. It will first focus on the necessity of underlying methods and proceed with an examination of the technical system that is constructed to put these into practice. Thereafter, hypothesis and variables will be linked to the research subquestions and the specific methods that will be used to analyze these are introduced. The scope, biases and privacy matters in exercising the proposed method will also be discussed in this chapter. Chapter four will present all relevant results and discuss their meaning, before findings will be put into a broader context in chapter five. Chapter six will combine all insights to answer the main research question and subquestions. This final chapter will also discuss the scientific and societal contributions that these insights bring. Further background material, such as the material that was used for communication and data analysis is added as an appendix to this thesis research.

2 Literature review

Following the introduction and goal of this thesis, this chapter will provide further background to the identified knowledge gap. The first section will introduce a history of privacy and privacy law in Europe, leading up to an account of contemporary theories in the modern literature on privacy and data privacy, in increasing detail. Further sections will focus on research into the practical application of this regulation, with specific focus on the execution of personal rights and effects of the 2016 General Data Protection Regulation (GDPR) on this execution. These backgrounds will combine in the last section to substantiate the knowledge gap that was previously introduced. The research question that is designed to fill this gap will be introduced and divided into subquestions that will guide the research design in chapter 3.

2.1 Legal history

The definition of privacy has been the topic of philosophical and legal debate for centuries, eventually leading to the introduction of foundational privacy rights in European law in the early 19th century. But even after agreeing on legal definitions, the interpretation and scope of privacy rights were often challenged because of changing social climates, organizational developments and technological innovations. These influences are also present in today's discussions on informational privacy, concerning the use and storage of personally identifiable or otherwise sensitive information on individuals (Holvast, 2009; Lukács, n.d.; A. F. Westin, 2003). In explaining modern definitions of privacy and the discussion that surrounds them, it is important to investigate earlier discussions that have us here. (Holvast, 2009; A. F. Westin, 2003)

The following section will describe the road of modern privacy definitions and regulations in Europe, therein analyzing the social, political and technological factors that ultimately led to the GDPR regulation. This analysis is loosely based on the conceptual framework used by Westin (2003). It should be noted however, that Westin's his paper describes development in the USA. Different social, technological and political factors have led Europe to a different direction in both the regulation and definition of privacy.

2.1.1 Historical context

The desire for privacy is said to be as old as mankind, (Lukács, n.d.) this idea that privacy is a fundamental human instinct is also shown by existing tribal societies and their wish for personal confined spaces for sexual privacy (Beach & Diamond, 1977; Ford & Beach, 1951). The classic Greek philosopher Aristotle described this innate need for a private domain as a distinction between two spheres, a public sphere containing politics and political activity named 'polis', and a private sphere called 'oikos' which included the domestic and family sphere of life (Beach & Diamond, 1977; DeCew, 2018). This separation of two domains is also found in the first known legal definition of privacy. In the early 17th century it was argued that "The house of every one is to him as his castle and fortress" in the English common law (Blakey, 1964). A more general definition of privacy following this reasoning, was formulated around 1890 as: "*a right to be left alone*" (Warren & Brandeis, 1890). This formulation still underpins of privacy regulation in the United States of America (Skousen, 2002)

European ideas concerning privacy diverged from those in the US after a baseline period following the second world war. During these first post-war years, western countries wanted to reaffirm the four freedoms that allied nations fought to restore in Europe: the freedom of speech, freedom of worship, freedom from want and freedom from fear. (Roosevelt, 1941) This led to the 1948 UN Universal declaration of human rights, which marks the first international statement on the right to privacy: *"No one shall be subjected to arbitrary interference with his privacy, [...]. Everyone has the right to protection of the law against such interference or attacks"* (Assembly, 1948; Morsink, 1999). This declaration, and a wish to counterbalance the rise of

communism in Central and Eastern Europe spurred the 1950 European convention of human rights (ECHR) to reaffirm this statement and establish the European Court of Human Rights to protect human rights (Frowein, n.d.; Ovey & White, 2006). Privacy was explicitly included in Article 8, “Right to respect for private and family life”. (Council of Europe, 1950)

Discussions on informational privacy, the relationship between data collection, processing and privacy, were not absent from public life at this time (Michael & Michael, 2014). This is best shown by the highly influential dystopian novel “Nineteen Eighty-Four” by George Orwell that was published in 1949 and describes a government that uses technology to control, censor and brainwash its people (Owell, 1949). However, developments in information technology were still limited and trust in western values and governments was relatively high because of the remarkable post war economic growth (Inglehart, 2008). Broader public discussions started around 1960, when the ugly parts of the second world war, such as the widespread Jewish persecution became increasingly open for discussion in western European countries. The availability of a proper population database and compulsory and incorruptible identification that was available in European countries such as the Netherlands proved to be deadly effective tools in the Nazi-government's ethnic cleansing goals. (Overkleeft-Verburg, n.d.)

This historical awareness, combined with the polarizing events such as the Vietnam and Cold War provided a big social influence on people's views on personal data. The increasing potential of automatic processing using computers, and their necessity to efficiently administer the growing welfare states of western Europe further strengthened concerns for citizens privacy. (Haren, 2017; Waxman, 2018) Meanwhile, inhabitants of Eastern European countries experienced the dark side of their government’s increasing technical capabilities as practiced by a security apparatus that did not care for their privacy. Practices of the Eastern German Stasi secret police further strengthened the post-Nazi feeling of significance of privacy rights in Western Germany, leading to the first European modern data privacy legislation being enacted in 1970. (Waxman, 2018)

2.1.2 Legal framework

Since the second world war, the European understanding of privacy has changed in response to technological, social and political forces. As opposed to narrower early definitions, the concept of privacy now also includes informational or data privacy (Michael & Michael, 2014). In the European Union, the legal framework that formed to protect informational privacy is driven in large part by the concept of people’s informational self-determination, obligation to lawful and safe processing for the party controlling the data processing (the controller) and the facilitation of international trade. Ever changing social, political and technological factors have further encourage the creation of increasingly specific, harmonized, enforceable and international regulation.

By 1979, specific privacy laws had been enacted in 9 European states, with more states in the process of implementation. (Faradina, 2017). Differences were abundant though, with some of countries focused more on (personal) privacy rights, while other countries’ legislation focused more on the facilitation of trade. (Charlesworth, n.d.) The Council of Europe became concerned that divergent national legislation did not adequately protect the populations privacy, especially considering further technological progress. To protect article 8 of the 1950 ECHR, European recommendations and resolutions were created to protect informational privacy with respect to new technologies such as automated data banks. The council also cooperated with the OECD and European Community to create influential guidelines concerning cross border data flows and the harmonization of national data protection law. These guidelines aimed to strike a balance between privacy and personal rights without creating trade barriers between countries. Many of the principles that are set out in the guidelines still resonate in modern European privacy law. And, although these guidelines are not binding, it did set the tone for the Council of Europe’s convention 108, which in 1981 was the first legally binding international data protection instrument (Rudgard, n.d.)**.** It states that:

- Personal data is defined as information relating to an identified or identifiable individual. This individual is further defined as the data subject and the party that decides about the storage, collection, processing or dissemination of the subject's data.
- Personal data undergoing processing should be:
    - processed fairly and lawfully,
    - stored for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes;
    - the data collected is adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;
    - that it is accurate and, where necessary, kept up to date;
    - that the data is kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed.
- Processing special categories of data such as race or political beliefs is prohibited
- Regulation can be derogated for state security, public safety, monetary interest or criminal offences and to protect the rights and freedoms of others.
- Any person should also be able to:
    - Establish the existence of personal data file and the identity of the controller
    - Obtain confirmation whether personal data are stored and received this data in an intelligible form
    - Obtain rectification or erasure if data has been processed unlawfully

(M. D. Birnhack, 2008; Council of Europe, 1950)

The 1980s mark a second era of privacy development. Further technological developments in this era do not only give governments more capabilities to investigate the personal lives of citizens, but also increasingly provide businesses with these capabilities (Privacy Europe, 2018). And although earlier European conventions, guidelines and resolutions all aimed to harmonize data protection approaches, even the legally binding convention 108 did not achieve this goal. A diverse set of legislation was developing even among countries adopting laws based on the convention. One example of this is the influential 1983 ruling by Germany's highest court that citizens have a basic right to self-determination over their personal data. This means that individuals should in principle have the right to determine (limits to) the use of his/her personal data and are protected against unlimited collection, storage, use and disclosure. (Hornung & Schnabel, 2009).

To counteract diverging regulation and strengthen the internal market, the European commission set out to create a more harmonized regulation, which would ultimately become the 1995 Data Protection Directive (DPD). In harmonizing regulation among member states, it built upon the framework of convention 108 and generally strengthen passages towards the strictest member state interpretation. (Swire, Litan, & Litan, 1998). The interpretation by German courts that privacy rights included the right of informational self-determination was thus included in the new directive. The convention's statement that a person was to obtain basic information on the processing of his personal data was thus transformed to the following fundamental data subject rights:

- Right to information/notification when personal data is collected
- Right to access the subject's personal data, in intelligible form and including knowledge of the involved logic.
- Right to object to the processing of personal data
- Freedom from automated decision making, giving subjects the right to object to decisions made without human interference.

Exceptions on these rights were also introduces: for national or public security, defense, criminal offences, important economic and financial interests of member states or to protect the rights and freedoms of others. Other key changes of the DPD with respect to convention 108 include:

- Stricter limitation on personal data processing. In addition to the principles of convention 108, personal data may be processed only under certain circumstances:
  - if the data subject has unambiguously consented to the processing;
  - if it is necessary for performing a contract to which the data subject is a party or complying with a legal obligation,
  - to protect vital interests of the data subject
  - if processing is necessary for the public interest or for the controller's legitimate interests, if these are not overridden by the subject's fundamental rights and freedoms.
- Member states should choose what special categories of personal data are prohibited.
- Member states are obliged to establish a national Data Protection Authority (DPA). Data controllers are obliged to maintain the confidentiality and security of data and to notify this national authority before carrying out automatic processing operation. The supervisor should carry out prior check to determine data processing risks and publish the notifications in an open register.
- Transfer of data to third countries is only permitted if an adequate level of protection is ensured, or if unambiguous consent is given by the data subject.

Exceptions on this stricter regulation was also introduced for cases such as processing for journalism, research, freedom of expression or healthcare.
(M. D. Birnhack, 2008; Directive 95/46/EC, 1995)

As the main effort of this directive was to harmonize regulation among member states, it did not radically depart from already existing laws in most big European economies. (Swire et al., 1998). It's prohibition of data transfer to outside countries without adequate measures did however have an impact on countries outside of the Union, as these risked being excluded to data flows with the whole European economy if adequate protection levels were not met. A committees of member state representatives was established to help the European Commission assess the adequacy of countries' regulation. The committee was later supplemented by representatives of national DPA's and the European Data Protection Supervisor (EDPS) (EDPS, n.d.).

Before the data processing directive was succeeded by further laws, it was seen as the only effective, comprehensive and successful international instrument of data protection law and the leading force of globalizing data protection. (Bu-Pasha, 2017) However, it has also received a lot of critique since its implementation in practice. These critiques can be summed-up in four categories.

1. Diverging regulation.

Since the DPD was an EU directive, the legal act was not directly legally binding regulation. Rather, an EU directive states the envisioned results that member states should achieve, without dictating the means of achieving this result. Member states are thus left with leeway to implement their own regulation. (Folsom R., Lake, R. B., Nanda, V. P., 1996) This meant that EU data privacy regulation was still not fully harmonized between European countries. Differences in data privacy regulation among members states further increased over time as extra laws where introduced in member states to deal with specific issues such as governmental databases and healthcare documentation. This divergence made it increasingly challenging for organizations to comply to both the specific national regulation and the general Data Privacy Directive. (Custers, Dechesne, Sears, Tani, & van der Hof, 2018; Zwenne et al., 2007)

2. Weak enforcement

Because of low DPA capacities and small fines, organizations often do not feel obliged to follow the rules set out in the DPD. Without the tools or resources to force compliance, their regulatory authority under the directive is often compared to that of a paper tiger. Without strict enforcement, many privacy rights that seem strict on paper, fell into obscurity (Seiler, 2016). Most European citizens do not know their personal privacy rights or are unable to do so because of disregard for these rights both by private and public data controllers. (Ausloos & Dewitte, 2018; Schoneveld, Spanninga, Sprenger & Postma, 2017; Norris et al., 2017) Even judges can be unfamiliar with these obscure rights, leading to different interpretations on their scope (Zwenne et al., 2007).

3. Modern technology

Some parts of the directive, such as obligatory notifications for every data processing case including personal data are not in line with modern use of technology (Zwenne et al., 2007). The development of tools that process huge amounts of information and the interconnected-ness of the internet have also introduced new business practices such as joint processing of data that are not addressed by the DPD, and thus hard to regulate using this old directive. Some critics believe that this is a symptom of fundamental flaws in the DPD's fundamental concepts, claiming that ideals such as informational self-determination are impossible to achieve in the reality of the 21st century economy (Koops, 2014).

4. Internationalization

The DPD is also seen as fundamentally outdated with respect to the modern, internationally connected world (Robinson, Graux, Botterman, & Valeri, 2009; Zwenne et al., 2007). Basic concepts such as accountability and the lawfulness of internal data transmission are unclear, cumbersome or not defined for multinational corporations. (Zwenne et al., 2007)

It is interesting to see that these four critiques are very much in line with critiques that were levied against earlier privacy regulation. Internationalization, enforceable protection of human rights and addressing technological innovation are the leading reasons to improve the DPD that was designed to guarantee exactly these concepts.

Some smaller European regulation has changed since the introduction of the 1995 directive. The 2002 ePrivacy Directive (ePD) and the 2007 Lisbon treaty, for example, did (attempt to) update parts of the regulatory framework surrounding data privacy. These smaller regulatory additions, however, did not address the previously introduced critiques of the data protection directive. These would have to wait for a total renovation of EU informational privacy regulation in the 2010's. (Machanavajjhala, Kifer, Abowd, Gehrke, & Vilhuber, 2008)

2.1.3 The GDPR

The process of the latest big update to European informational privacy legislation, which eventually became the General Data Protection Regulation (GDPR), was created by the EU Commissioner Viviane Reding in November 2010. In her publication addressed to all EU legislative bodies she started the formal process of reforming the 1995 Data Protection Directive that, "set a milestone in data protection by enshrining the important ambitions of protecting fundamental rights and creating an internal market" (Reding, 2010).

After public consultation and multiple studies, Reding's publications points to five key issues with regards to EU directive 95:

1. The impact of new technologies, which were not properly addressed in the directive.
2. Enhancement of the internal market, by harmonizing member states' legislation.
3. Globalization and international data transfer, which made application of the directive challenging.
4. Stronger institutional arrangements, as enforcement by DPA's was uncommon.
5. Improving legal framework coherence, with overarching instrumentation.

(Reding, 2011)

These five issues contain most of the earlier referenced critique on the DPD, but also clearly match the motivations that were once expressed for its creation.

What followed was long period of drafts and redrafts, as the European Parliament, Council and Commission negotiated a joint proposal. In this process, the commission and parliament specifically called the strengthening of citizens' rights as one of the primary rationales for reform. The Commission composed of appointees from each EU country was also focused on the international trade and business aspects of data privacy law. When the commission's first draft was presented to the parliament and council in 2011, 305 external written responses and further targeted consultations were already included. The European Parliament's advisory committees proposed a total of 3133 amendments to this text and published its version in 2014, with the European council creating its position on the regulation with a different document in 2015. (Norris et al., 2017)

As the European commission and council did not agree with all amendments proposed by the Parliament, a conciliation committee was appointed. This committee was tasked with reaching a joint text that will be accepted the three legislative bodies. This joint document was finalized in December 2015 and accepted in parliament on April 2016. Two years after the official publication, the GDPR has come into effect on the 25th of May 2018. (Wilhelm, n.d.; EDPS, n.d.b; Norris et al., 2017)

It is no surprise that this lengthy process, extensive disagreements and the potentially far reaching consequences have motivated a lot of other parties to influence the decision-making process (McNamee, 2015). In fact, the GDPR is said to be the most-lobbied piece of European legislation ever written. (Trevor Hughes, 2016) Many business groups, privacy advocates and other interest groups (including governmental parties such as the US chamber of commerce) tried to tilt the legislation in their direction (Norris et al., 2017). Critics have pointed out that some of these lobbying groups, especially those acting on behalf of business interest have indeed managed to reword key passages in further GDPR drafts. These groups, on their part point to other passages in the final draft that are (still) way to strict in their opinion. (Singer, 2018; Financial times, 2017)

Still, the outcome is considered by many to be a new European milestone in privacy regulation (Trevor Hughes, 2016). One of the most important changes is that the General Data Protection Regulation is legally binding in all member states, as opposed to the 1995 directive which was separately translated by each member state into national law. A full overview of substantive changes to the new regulation with respect to the DPD is included in Appendix A.1.

As the review shows, the GDPR introduces or strengthens a lot of different obligations for data controllers while only simplifying one (the obligation to report all personal data processing to the DPA). In response to new business practices, a distinction is made between data controllers' and processors' rights (when data processing is outsourced by a data controller) and obligations. It also increases the DPA's rights and obligations and increases the territorial scope of the regulation. With these changes, it is clear that the regulation is stricter than its predecessor. This is further exemplified by a renewed focus on people's individual rights. The European Union's regulatory bodies are thus convinced that there is a place for guaranteed informational self-determination rights the 21st century. The second Table in Appendix A.1 describes specific changes to these rights of Access, Rectification, Cancellation and Opposition of the processing of personal data. Out of these ARCO rights the right of access is the 'natural precondition' for the other rights, as data subjects first need to know what personal data is processed and by whom before being able to practice the other rights (L'Hoiry & Norris, 2015). Table 2 shows changes that the GDPR brings to this right in particular.

Much of the critique launched at data privacy rights regulation in the previous section has been addressed at some level in the new set of regulation. It is clear that the GDPR, at least on paper, presents more solid safeguards to the right of access and data privacy in general. The GDPR also presents a new focus on ARCO rights, with personal privacy being defended stricter than ever before. The following section will investigate the translation between the written law and its application in practice.

Changes to the right of access, GDPR art. 12 compared to DPD
Specified time limit of responses from "without […] excessive delay" to within one month (with the possibility to extend that period with two more months when necessary but then explain why)
Specified that these rights may only be restricted through union or member state law when these restrictions are specific, respect the essence of the fundamental rights and freedoms, is necessary and proportionate in a democratic society.
Specified exceptions to the right of access for only scientific or historical research purposes (only if appropriate safeguards are met such as data minimization and
Added exceptions for data processed for only achieving purposes in the public interest or statistical purposes "if access requests are likely to render impossible or seriously impair the achievements of this purpose".
Added controller obligation to:
• Provide the right of access for free (Except when requests are manifestly unfounded or excessive, or if further copies are requested)
• Added controller obligation to provide for means of requests to be made electronically (especially when personal data are processed by electronic means.
• Added controller obligation to provide requested data in writing or where appropriate, by electronic means. Where possible, via direct remote access to a secure system. (When requested, information may be provided orally too)
• Use all reasonable measures to confirm the identity of the subject for which a controller may request additional information.
• To report in a concise, transparent, intelligible and easily accessible form, using clear and plain language. (Especially when children are involved)
• Added obligation to disclose, when exercised:
• Not only concerns recipients or categories of recipients that are disclosed, but also those that will be disclosed.

Much more information on automated decision making, including its significance and the envisaged consequences.
The envisaged retention period or criteria used to determine that period (“where possible”)
The existence of the right to rectification, erasure, restriction, objection and to file complaints w/ DPO
The appropriate safeguards relating to transfer to data to a third country or international organization

Table 2. *Differences between the right of access as described in the GDPR with regards to earlier implementation in the DPD. Sources: (Faradina, 2017; Ausloos & Dewitte, 2018; Birds, 2017; Directive 95/46/EC, 1995; Regulation (EU) 2016/679, 2016)*

2.2 Privacy in practice

Ever since the idea of informational self-determination began to take hold in the European mindset, privacy has changed from a passive right to an active one. In the previous section we have seen that since 1981, European citizens have the right to “establish the existence of personal data”, “identify the controller” and “receive this data in an intelligible form”. These privacy rights have been increasingly prominent in later legislative overhauls, with special focus on the right of access, which acts as a natural precondition to the others (L’Hoiry & Norris, 2015). The new GDPR places even further emphasis on this natural precondition for transparency and informational privacy. It’s more stringent inclusion in written law, however, does not directly lead to an increase in the privacy in the life of citizens. This section will explore literature on the use of privacy rights in practice and how the GDPR will influence them.

2.2.1 Exercising privacy rights

Authors have written extensively about personal privacy rights since the second world war and the application of such rights into law are seen as a cornerstone of data subject’s empowerment. (Ausloos & Dewitte, 2018; Norris et al., 2017). Early writing on practicing these rights in practice has been scarce however, with most publications focusing on legal analysis instead. This created a situation in which only a small number of experts knew of the existence of access rights. The relative obscurity only made exercising privacy rights more challenging, which only made its practice rarer and thus more obscure. Before 2010, the right was almost exclusively used by lawyers and activists. (Winter et al., 2008) This state of right usage is in in sharp contrast with the original intent of the law and the public’s strong desire to have and be able to use access rights. (Grogan & McDonald, 2016)

The use of access requests returned to public interest in 2011, when Austrian law student Max Schrems requested a copy of his personal data that was processed by Facebook. The more than 1200 pages of data that was eventually received showed multiple breaches of European regulation which were pursued in various European jurisdictional levels. Schrems showed that the obscure right to make these Subject Access Requests (SAR’s) did not only enable citizens to gain insight into use of their personal data by organizations, but could also lead to legal actions, societal awareness and legislative reform. An increased European interest in informational self-determination was further fueled by the publications of the Snowden revelations and WikiLeaks documents. The widespread surveillance programs that were uncovered lead to the annulation of the EU-US safe harbor agreement increased public privacy awareness and decreasing public trust.

In academic literature, considerable research is published concerning surveillance and the social and ethical challenges that it brings. Concerning personal privacy regulation, the rights of notification, rectification, cancelation and the right to be forgotten also received scholarly interest. However, the “natural precondition”(L’Hoiry & Norris, 2015) in exercising these other rights, was in an investigatory vacuum until 2017 when leading privacy researchers published “The Unaccountable State of Surveillance” (Norris et al., 2017), which features quantitative and qualitative analysis of data subjects experiences in exercising their access rights in ten European countries. Although its conclusions were not surprising to those that had tried

to practice these rights before, it clearly showed that there was a big disparity between privacy rights in theory and practice:

- In trying to send access requests, researchers were unable to locate the data controller in an average of 20% of cases.
- When access requests could be submitted, only an average of 43% cases yielded a positive outcome (defined as responses that adequately addressed three researcher's queries and thus provided the legally required insight).
- Results varied greatly between countries (with 71% positive outcomes in the UK versus only 33% in Italy and Norway).
- When submitting complaints to national DPA's, only 64% of these were resolved

Researchers were able to receive the legally required insight in only 34% of cases. It was therefore concluded that *"many organizations [...] have clearly abrogated their legal responsibilities"* with a broader implication to policy maker that: "*unless there are mechanisms to ensure that the legislation that they enacted is transposed into the routine practices [...,] it is likely to be of marginal value"* (Norris et al., 2017). Following this ground-breaking insight into privacy law in practice, similar methods were used in research by others. Their results are summarized in Table 3.

Study	N	Country	Sectors	Response rate (1)	Response with data (2)	Response with answers (2)	Compliance
(Norris et al., 2017)	183	EU	Various	80%	57%	43%	34%
Asghari, Mahieu, & van Eeten, 2017	106	The Netherlands	Various	83%	69%	27%	22%
(Herrmann & Lindemann, 2016)	120	Germany	Popular apps & websites	68%	n.a.	n.a.	43%
(Spiller, 2016)	17	UK	CCTV	n.a.	35%	n.a.	35%
(Ausloos & Dewitte, 2018)	60	EU	Various	74%	67%	n.a.	(3) 33%

Table 3. *Previous results from research into responses by organizations to subject access requests.*
(1) The response rate includes responses which are received after the legal deadline and/or received after multiple reminders which can be considered non-compliant to the law. (2) Data that was deemed incomplete or inaccurate is excluded, if this was found in the results. Responses without data that were deemed accurate by participants are categorized as data provided. (3) Based on satisfaction rating by participants, as legal outcome was not reported.

Some of the studies in this Table expand on the results for Norris et al. (2017), by analyzing differences in responses by organizations in different sectors (Mahieu, Asghari, & van Eeten, 2017) or combining research into the rights of access with the right to be forgotten (Herrmann & Lindemann, 2016). However, none of the performed studies contradict the general conclusions that were put forward by Norris et al. (2017). The right of access has consistently failed to provide European citizens with information that was rightfully theirs in over 50% of cases and can therefore be considered a failing instrument. This conclusion is consistent with research into both public and expert opinions. These insights further fueled critics claims that the GDPR renewed focus on informational self-determination is unrealistic and outdated (Koops, 2014; Norris et al., 2017).

However, this is not a view that the authors express in their studies. Instead, multiple options are put forward to salvage the right of access as a privacy instrument. Some of these are present in the GDPR, such as the harmonization of access requests regulation across EU countries, the representational rights of organizations to help prosecute violations and the general increased hype around data protection that is expected to further inform data subjects in their rights under the GDPR (Norris et al., 2017). Other proposed changes are not

implemented, such as the removal of a business legitimate interest as a justification for lawful data processing, requesting 'explicit' over 'unambiguous' consent and refusing member states to add extra restrictions on access rights. (Ausloos & Dewitte, 2018; Norris et al., 2017) As described in the previous section, it is generally agreed that subject access rights under the GDPR are stronger than under the 1995 directive that was relevant for the results in Table 3. However, scholars disagree in their assessment if this influence is big enough to salvage the right as an effective tool to guarantee privacy. (Asghari et al., 2017)

2.2.2 The GDPR in practice

After the GDPR was approved by the appropriate legislative organs and published in the official EU legislative journal at the 25th of May 2016, it kicked off a two-year period before becoming enforceable. This period was designed to provide organizations enough time to adjust their processes to the new regulation. In practice however, most organizations delayed implementing changes until the 2018 deadline was unreachable (Chiavetta, 2017). When this deadline was near, many organizations, organizations, governments and even enforcement agencies were far from compliant (Curtis, n.d.; Ponemon institute, 2018). The potential risks associated with non-compliance fueled a climate of panic around the GDPR, which was only further agitated by those that profited by providing (costly) consultancy services or legal advice on the matter. (Cellan-Jones, 2018; Ustaran, 2017). This panic reached its peak towards the 25th of May compliance deadline (Dots, 2018), when data subjects received huge numbers of emails by organizations asking data subjects to (re-)consent to data processing. These emails were often unnecessary and sometimes illegal forms of spam under the 2002 e-privacy directive and may have left consumers with a bad sentiment towards the GDPR. (Hern, 2018)

Large tech organizations such as Google, Instagram, WhatsApp and Facebook had lawsuits filed against them immediately following the deadline for violating the GDPR by "forcing users to agree to new privacy policies". (Denhart, 2018; noyb.eu, n.d.) The enforceability date of the GDPR also lead to a steep drop in demands for targeted online adds in the European advertising market (although it has started to recover since) (Davies, 2018; Joseph, 2018). The new regulation is also blamed for a small drop in the number of active users on some social media platforms, although this might have other causes. (Lanxon & Bodoni, 2018). A small number of international organizations decided to avoid risks by not servicing European citizens altogether. This includes online marketing firms (Schiff, 2018), smart appliances (McMullan, n.d.), online tools and US local newspapers (Hatmaker, 2018). At the time of writing, three months after the GDPR came into force, most of these services are still not available in the EU. A small number of services have decided to offer different functionalities or prices for European users to make up for the burden of GDPR compliance (Rijo, 2018; Statt, 2018).

Although the exclusion of services and organizations to the European market can lead to decreased innovation and competition, the negative effects of GDPR enforceability have not been as apocalyptic as predicted (Kottasová, 2018). Furthermore, although the hysteria that surrounded the deadline certainly did harm certain organizations (Hugo Monteiro, 2018a, 2018b), it also showed that the new GDPR did get organizations to reassess their use of personal data, thereby dispelling the myth of an unenforceable law which organizations would never try to comply with. This is especially true for non-EU multinationals, which often reacted strongly to this regulation. Their reaction is explained as a manifestation of the Brussels effect which shows the global power of the European Union in exporting legal influence through market effects (Bradford, 2012). Some international organizations, such as Microsoft and Sonos have announced global implementation of the GDPR (Brill, 2018; Sonos, 2018). The US state of California has already voted to implement a similar bill (State of Callifornia, 2018), with other countries planning to follow suit (Scott & Cerulus, 2018).

2.3 Knowledge gap

The buff of personal access rights in the GDPR is prominently present in most guides to GDPR compliance, which can be found all over literature (Gibbons, 2017; Fox, 2018; D. Jones, 2018; Karbaliotis, 2017). Even after the GDPR has come into effect, the tools and practices that organizations need to implement seems to still be the dominant topic of research. Research that does investigate the implications of the GDPR so far has focused on the written law instead of the implications in practice (Davis & Osoba, 2018).

An inquiry into sources outside of academic literature paints an inconclusive, but overall improved idea of subject access rights past the GDPR. The number of filed Subject Access Requests (SARs) seems to have increased in certain countries and sectors (Fiore, n.d.; Hill, 2018), which seems to imply that the tool is more widely known. DPA's are also better known, with a larger number of complaints being filed after enforcement of the GDPR. (Autoriteit Persoonsgegevens, 2018; Press Association, n.d.). Previously available online tools, such as the Bits of Freedoms Privacy Insight Machine (PIM) are being upgraded to better support data subjects in their requests (BOF, n.d.; My Data Done Right, n.d.), while others are being designed from the ground up for the same purpose. (datarights.me, n.d.; Dehaye, n.d.; Dehaye, Hahn, & Jargalsaikhan, n.d.) Tools and programs that help organizations with proper responses to these requests are also being created. (Clarip, n.d.; Sesam, n.d.; TrustArc, n.d.). The implementation of these tools should improve the process of exercising privacy rights. However, their effect on this process itself remains unclear.

The only recorded instance of subject's access rights in practice under the GDPR known to the researcher is a Dutch newspaper article detailing the responses of ten big organizations that were asked to give insight into the personal data of two journalists. (Verhagen, 2018) Although the article is a great example of a potential use of access rights for both personal and societal benefits, it's small sample size and inconclusive results do not lend itself for generalization. The lack of large-scale quantitative studies into the application of a novel law is not surprising, but still disappointing given the revelations that this type of research brought to the state of privacy under the 1995 privacy directive.

After years of discussion, the European Union choose to include a more uniform and strict application of the idea of data access rights, thereby stating its renewed intention to guarantee its citizens informational self-determination. It is however, woefully unclear if these changes have really delivered the informational rights that privacy law has tried to guarantee since 1983. This research will aim to fill this knowledge gap, thereby answering the following research question:

How are organizations responding to Subject Access requests under the GDPR,
and what factors influence their responses?

The answer to this question should not only lead us to a description of the current state of affairs as influenced by the GDPR. Rather, changing regulation might only be one of the factors that influence organizations response to SARs. In fully answering the research question, this thesis aims to also investigate other factors that may influence this response.

2.3.1 Conceptual model

Before the next chapter can delve deeper into practical considerations that accompany an investigation into responses to access requests, it is important to investigate the factors that influence responses to these requests. A conceptual model that shows these factors and relations will help steer the research towards the measurement of relationships that answer our research question. Since no such model is found in relevant literature, one will be created in this section based on earlier research.

Just like under earlier law, the "Right of access by the data subject" under the GDPR is an active right. In a high-level summary, the process of exercising the right consists of three steps. First, the data subjects need to actively request a data controller for insight. When receiving this request, it is the controller's obligation to check if the request is valid and if so, investigate what personal data is being processed by the controller (or on behalf of the controller). When requested, the controller should also investigate further details on how this data is processed, like with whom it is shared and how long it is stored. When this internal process is finished, the controller will send its findings to the data subjects. These findings can be a ground for full or partial denial of the request and/or a certain degree of insight into the processing of the subject's personal data that is carried out by the controller. An adequate response means the access right process is finished, although a data subject might repeat the steps to receive further insight.

The approach on how these three steps should be undertaken is generally described in article 15 of the GDPR. Best practices and examples of the first step can be easily found online, for example on the website of national DPA's. (Recht op inzage, n.d., Right of access, 2018a, voorbeeldbrief verzoek om inzage.pdf, n.d.). Advices on how to perform the second and third step are even more numerous, with a wide range of freely available literature by DPA's, scholars and professionals describing practices that should be followed (Bijron, 2017; Inventis, n.d.). An even larger collection of advice and (best) practices can be obtained through payment by either buying specialized tools (Urglavitch, 2017), hiring experts (The EU GDPR, n.d.) or outsourcing the GDPR compliance process altogether (Shepherd, n.d.).

The wide availability of advice on internal processes does not guarantee correct processing of requests within an organization however. Furthermore, the existence of an internal process prescribing proper conduct does not guarantee its use in practice. Rather, organizations execution in practice is bound by resource restrictions in and incentives to allocate these resources in correctly processing access requests. (Asghari et al., 2017) Earlier research on exercising access rights in practice is therefore always performed from the perspective of the data subject.

From this perspective, the second step in the SAR process (processing of this request by the controller) is hidden to the subject, even when the internal processes that are theoretically being followed by the processor are available. The subject only observes the input (submitting the request) and output (return of findings or other responses) of a black box (Cauer, Mathis & Pauli, 2008). One often repeated flaw of access requests comes from this inability of the data subject to judge the process that lead to the response, since on individual does not know if the data controller indeed included all of the relevant personal data in the response. (Koops, 2014)

A larger set of responses however, can help judge the response of certain organizations or groups of organizations through comparison. This collective endeavor "*can help shift the power imbalance between individual citizens and organisations in favour of the citizen, which may incentivise organisations to deal with data in a more transparent way*" (Mahieu et al., 2017). Furthermore, trends that are observed in a larger number of organizations may offer a glance into the organization's black box process by identifying variables that effect certain types of responses.

A model of factors that may influence the execution of access requests by data controllers on a societal level is constructed by looking at research into regulatory compliance in other fields, disparities found in results of earlier practical studies and assumptions based on known organization practices. The following six categories of factors are believed to play a role in an organization's responses to subject access requests:

- Request factors. What request is sent to the data controller should influence its response. This does not only include the content of the request, but also the tone and the transmission method. Especially this last factor is of importance in this research, for reasons that are explained in chapter 3.

- Judicial factors. This category covers all judicial differences that might influence an organization. For example, what is stated in current regulation that the organization should follow, what repercussions can be expected and what chance an organization has to receive these by not following the law. (Ausloos & Dewitte, 2018; Norris et al., 2017)
- Organizational factors, encompassing differences between organizations. organizations operating in different sectors seem to react differently to the same access request (Asghari et al., 2017). Organizations of different size, location and age might also react differently to these requests, because of differences in capabilities or priority given to access requests.
- Subjects factors. Organizations' responses to the same access request might also differ because of a difference in the subject that requests this response. This might be because the request is sent by a different person, for example by a lawyer of citizen group on behalf of the data subject, or because of a different relationship between the subject and the organization (e.g. customer, employee, business relation).

The relationship of these factors in the access request process is shown in figure 1.

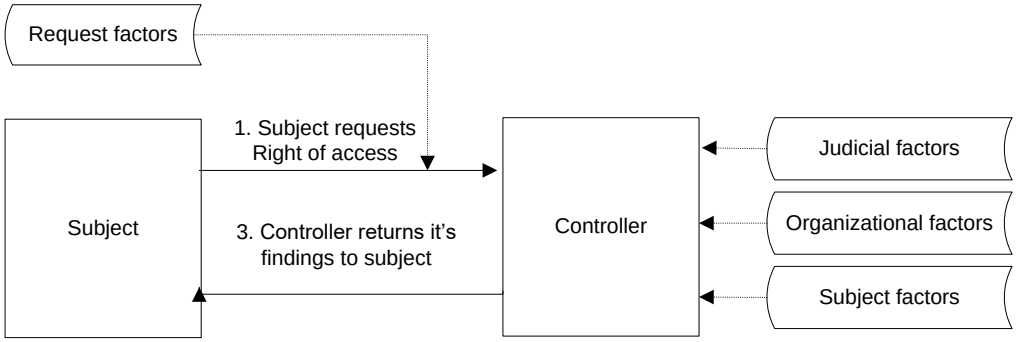


Figure 1 Conceptual model of the actors, process steps and factors influencing the process of subject access requests

2.3.2 Research questions

The way organizations react to subject access requests under the GDPR, is explained by the relation between the data subject's request and the controller's response. This relation can be explained on a societal level by investigating the effect that certain judicial, organizational and subject factors have on the data controller's process that produces this response. To answer the research question put forward in the previous section:

How are organizations responding to Subject Access Requests under the GDPR,
and what factors influence these responses?

research should thus aim to answer the following subquestions:

1. *What judicial factors influence data controllers' responses to subject's access request, and to what extend?*
2. *What organizational factors influence data controllers' responses to subject's access request, and to what extend?*
3. *What subject factors influence data controllers' responses to subject's access request, and to what extend?*
4. *What request factors influence data controllers' responses to subject's access request, and to what extend?*

The following chapter will further elaborate on the way these relations can be observed, and what research is performed to measure their effect.

3 Methodology

Previous chapters have detailed the necessity of research into the way organizations respond to Subject Access Requests (SAR's) under the GDPR and introduced the research question and subquestions that should be answered in order to provide insight into these responses. The following chapter will feature the methodology that was designed to do just this. The first subsection will explain the approach and thought process that lead to designing research in this way. Section 2 and 3 will provide detailed information on the data collection and data analysis phase of the research respectively. The final sections will describe expected bias that accompanies this research, and the verification and validation steps to prevent tainting the results with these and other biases.

3.1 Research approach

In this research, a mix of qualitative and quantitative methods should be used to answer the research questions presented in the previous chapter. The quantitative part would be the best fit for investigating the effects of research questions 1 to 4, although a qualitative answer would also contribute to the scientific understanding of the framework of access requests. The exploitative nature of question 5 lends itself more towards an exploratory qualitative analysis, although quantitative results would be welcome additions to the answer. The mixed methods approach is most apparent in the main research question, where both the effect of certain factors and their context should be part of an answer. These methods are combined in a research approach that can be classified as an exploratory mixed methods approach.

To answer the research questions that were formulated in the second chapter, a dataset is needed in which responses to subject access requests can be connected to data controllers experiencing the different effects shown in the conceptual model of section 2.3.1. Qualitative analysis requires detailed results of 20 to 50 cases. (Creswell, 1998; Glaser & Strauss, 1967; Morse, 1994) Envisioned data analysis techniques such as logistical regression statistical testing generally perform better if more data is applied, but a rule of thumb lower limit is twenty cases per predictor (Urdan, 2016). With a heterogeneity of around 20 measurements per investigated indicator, the size of the envisioned dataset as a whole should contain between 100 and 300 measurements.

Dr. Asghari has pledged his help in providing an adequate number of accurate measurements of a range of organizations responding to access requests under previous DPD jurisdiction. This dataset, which was used for the research in Asghari et al. (2017) will be used to investigate the effect of the judicial change to the GDPR. However, a dataset containing such responses under the GDPR does not yet exist. Because this data is essential to answer subquestion 2 and the main research question, it should be collected for this purpose. The following section will detail the method by which it is performed.

After the GDPR response data is collected and combined with the older dataset, it should be classified to prepare it for data analysis. The variables that should be classified are operationalized from the in- and output factors in the conceptual model in section 3.4. This section will also detail how organization's responses are recoded into the ordinal results that are reported in chapter 4. Differences in this (ordinal) result variable will then be tested between different groups of organizations in order to investigate hypothesized effects. These tests will be performed using statistical techniques such as Mann-Whitney U, Wilcoxon and chi^2 tests and linear logistical regression models (Lalla, 2017) using jupyter notebook, a web-based application that provides an easy overview of data analysis code and its resulting output. When certain analysis tools are not available (as packages) for the default Python implementation of these notebooks, the R programming language will be

used. Quantitative results that follow from these statistical tests can then be supplemented by a qualitative analysis into the both the process and results following from SARs.

3.2 Data collection

Earlier research into responses to access requests under the Data Protection Directive provides a blueprint for creating such a dataset. Norris et al (2017), Asghari et al. (2017) and Ausloos & Dewitte (2018) all created their own dataset for similar data under previous regulation by requesting insight into the processing of their own personal data. However, applying this method directly to the research at hand creates a conflict with additional requirements. This problem arises from the practicalities of a thesis research that is undertaken by a single researcher within a timeframe of 6 months.

These constraints are in sharp contrast to previous research examples of Table 3, which often took multiple years to accomplish, were performed by multiple experienced researchers and often received further assistance from students (Ausloos & Dewitte, 2018; Herrmann & Lindemann, 2016; Norris et al., 2017). It is unlikely that a dataset which is larger and more detailed than what was used for earlier research can be collected in a shorter time span, by fewer researchers using the same method. It is therefore imperative that previous used methods are supplemented by a method that enables faster data collection.

This opportunity is found in the concept of distributed data collection, as described by Salganik (2017) in his book on the mass collaboration opportunities that the 'digital age' brings for social research. A good example of this collection method's capabilities is the PhotoCity project, in which researchers enlisted the help of citizens in providing the necessary 2D photos to create high-resolution 3D models of buildings (Tuite, Snavely, Hsiao, Tabing, & Popovic, 2011). The example shows three key point that make this method valuable for this thesis' particular purpose. First, enlisting the help of a lot of people created a larger dataset in little time and facilitated the collection of that was out of reach without these people (special angles, pictures certain times a day). Second, researchers could steer people towards collecting the most valuable datapoint, in this example towards the missing 2D pieces to create a complete 3D model. The third point is that collected data could still be validated by researchers before it was used for creating the models. Mass collaboration can thus enable the faster collection of a larger, more relevant dataset, while still enabling quality checks.

The application of mass collaborative data collection seems viable for Subject Access Requests (SAR). Third party tools that help people in their quest for access do already exist, and are expected to increase with the introduction of the GDPR (Ruiz, Johnson-Williams, 2018). And although these tools are often only used to create letters that the data subject then sends to the controller, the right of access is known to be used on behalf of others. This is often a solicitor such as an attorney requesting access on behalf of his client. In the previous chapter, no changes to the GDPR are identified that limited the data subject right to request data access through a third party. This is also highlighted by the ICO, the DPO regulatory body of the United Kingdom, on their website:

"The GDPR does not prevent an individual making a subject access request via a third party. Often, this will be a solicitor acting on behalf of a client, but it could simply be that an individual feels comfortable allowing someone else to act for them. In these cases, you need to be satisfied that the third party making the request is entitled to act on behalf of the individual, but it is the third party's responsibility to provide evidence of this entitlement. This might be a written authority to make the request or it might be a more general power of attorney." ("Right of access," 2018b)

Although the Dutch DPA did not explicitly state the legality of this method directly[1], Dutch property law specifically states that civilians have the right to declare a representative to perform legal action on their behalf. (Burgerlijk Wetboek Boek 3 - BWBR0005291, 1992) Comparative laws for exercising legal action through intermediaries can be found in other EU countries, although some countries require the declaration to be notarized (Dine, Koutsias, & Blecher, 2007; The powers of attorney, 2016). The notarization of intermediary statements is not required under Dutch law (Busch, H Hondius, J Van Kooten, N Schelhaas, & M Schrama, 2003). Collaborative data collection through an intermediary is therefore considered to be feasible and expected to facilitate accelerated collection of larger and more focused datasets. Please see the following sections for ethical considerations regarding this method and how the method is applied in practice respectively.

3.2.1 Ethical considerations

In using an intermediary approach to exercise data rights together, the data subject and researcher collaborate to achieve their respective goals. For data subjects, the researcher's assistance will make requesting data more accessible as it will enable even those without any knowledge of the relevant regulation to exercise their rights. The researcher's experience in the process of SAR's and the use of standardized tools will also ease the burden on data subjects during the process. This experience is particularly beneficial in dealing with unsatisfactory response from organizations. Furthermore, the data subject can receive this assistance for free, as the researcher to also be invested in the outcome. For the researcher, collaboration provides an opportunity to steer the data collection towards topics that are relevant for research. Being more closely involved, it also provides the researcher with more details of the process that could be lost when only the result is used for analysis purposes. Most importantly, this collaborative method extents the scope of organizations that can be used for research without recruiting more researchers.

Of course, performing access requests on someone else's behalf to use this data for research purposes provides a wide range of ethical challenges. The approach to these challenges was finalized with helpful advice from experienced researchers in multiple EU countries and ethics experts from Delft University of Technology. The approach was approved by the Deft University Human Research Ethics Committee (HREC) before data collection was started. It combines the principles of security and privacy by design, anonymization, informed consent and the freedom to withdraw this consent at any time without implications. The committee's application, which is added as appendix A.2 to this report, describes the applications of this concepts in detail. The most relevant choices are also present in the following section, which details the data collection practice.

3.2.2 Data collection system

After discussing the methods that are used in this thesis research and the ethical considerations that accompany them, this section will present the way in which the methods are combined in practice. Because of its novelty, the first two subsections will describe aspects of the applied data collection method, with a third section describing the research scope that follows from its application.

This research features a novel data collection method which often requires confidential communication between organizations, researchers and volunteers. As this communication is often carried out online, an online system should be implemented that allows researchers and volunteers to easily send (or reply to) requests and share responses between them. These requirements are not supported by current access request platforms. One was thus created specifically for this research using the tools that were available to the researchers.

[1] After data collection for this research was finished, the Dutch DPA has stated the legality of performing access requests on behalf of others on their website. See: https://autoriteitpersoonsgegevens.nl/nl/zelf-doen/privacyrechten/recht-op-inzage#mag-iemand-anders-voor-mij-een-verzoek-tot-inzage-doen-6780

The datarights.me domain lies at the core of the data collection system. This domain is hosted on a private TU Delft server and maintained by Dr. Asghari to allow safe and confidential communication and data collection for this research project. Volunteers have their own mail address and can access its content using roundcube, an open source webmail program that is hosted on the datarights.me domain. Using this program, researchers can send emails on behalf of the volunteer without accessing the volunteer's private inbox. A blind copy of emails that are sent on the volunteer's behalf will also be delivered in their respective inbox, to guarantee the volunteers knowledge of this action.

When an organization responds to an access request via email, this response will be delivered to the corresponding volunteer's mailbox. If the volunteer has indicated that a response from this organization may be shared immediately with the researcher, the response will be automatically delivered to the researcher's mailbox using Roundcube filter rules. If the volunteer has not chosen to automatically share replies from this organization with the researcher, he or she may still do so by forwarding (sections of) the email to the researchers email address.

Physical letters that are sent on the volunteer's behalf should also be shared with these volunteers. A digital copy is therefore made of each outgoing letter and emailed to the respective volunteer. Letters that are received by the researcher on the volunteer's behalf can be opened and answered by the researcher if he or she is authorized to read this communication by this volunteer and will otherwise stay unopened until authorized. When other communication methods are used, such as phone calls, personal conversation or webforms, the researchers should try to communicate the content of the communication to the volunteer. Organizations are asked to communicate using email whenever possible.

3.2.3 Volunteers

Because of the novelty of the applied method, data will be collected in two separate waves. The first group of 8 volunteers will ensure that at least an initial dataset can be created to answer the research questions when problems during further data collection arise. Experience from this test group can improve the process for the second group of 25 additional volunteers. The split-group collection method also provides researchers with a way to steer data collection to test hypothesis formed on results from the first group.

From the volunteer's perspective, the research will have three distinctive phases. Research starts with the inception phase, in which the researcher recruits and informs a new group of volunteers. Recruits should both provide meaningful additions to the organizations that can be investigated and be genuinely interested in the results of the subject access requests that are send (on their behalf). Because it is important that the volunteer understands all risks and choices associated with the research, the inception phase contains a face to face meeting of one hour with the researcher and volunteer.

Following the structure of the (Dutch) informed consent form that is included as an appendix A.3, volunteers are first informed of the research goal, context and division of roles. Subsequently, volunteers are asked to write down contact details required for the research, as well as some basic information that can be used to report the volunteer sample statistics. After this, volunteers are guided on a brainstorm to write down organizations that are expected to own their personal data. When an appropriate number of potentially interesting organizations is found, it is up to the volunteer to check if all organizations can indeed be used for this research and which organizations' responses can be shared immediately with the researcher for faster data collection. Only after deciding on all these details, the volunteer is asked to sign the research consent and communication consent form. In the end, a safe copy is made of the identity card of the respondent.

When all volunteers in a group have finished this inception phase, the researcher will decide which of the recorded organizations should be contacted on whose behalf. Since access requests may also require time investments of the volunteer, this number should be around 10-20 per volunteer. A final email containing these choices is then sent out to all volunteers, giving them a chance to block the inclusion of certain

organizations in case they have changed their mind. This marks the end of the inception phase, and the start of the communication phase. All access requests made to organizations are sent by the researcher using a standard request letter that can also be found in appendix A.4 of this report. This first request also contains a copy of the signed communication consent form and the proof of identity of both the researcher and volunteer and is sent using the communication method that is described in the organization's privacy statement. The communication phase continues with volunteers and researchers responding to organizations' replies to the request.

When data collection has reached a point where organizations have received appropriate time to correctly respond to the subject access request and no new responses are expected, the researcher will stop the collection phase. The GDPR states that organizations should respond to such requests within one month but may request an additional two months of response time when necessary. After the collection phase is finished, all volunteers in the group are invited for a second face-to-face meeting with the researcher. During this wrap-up phase, volunteers are asked to rate their experiences of the data collection method and subject access requests in general. The researcher and volunteer also go over the received data together. At this point, the volunteer can decide what information can be shared for further research.

3.3 Data analysis

After data is collected, it should be classified and analyzed to answer the research questions. This section will detail the process of linking research questions to specific criteria, creating hypothesis for these criteria and testing these hypotheses.

3.3.1 Classification

The classification of received data can be performed either by the researcher, or by the volunteer under guidance from the researcher, when the volunteer perceives the data as too personal to share. Organizations are asked to send their data in a machine-readable format whenever possible, which probably aids classification. Reclassification of the DPD dataset to the new variables is performed by Dr. Asghari, who also oversaw the original data collection, to minimize the loss of response details that could influence the classification. Either way, the organization's responses are classified using the variables in Table 4 below.

Variable	Criterium
Extension	Has the organization asked for an extension to the deadline to respond to the access request? (Y/N)
Data received	Did the organization provide personal data concerning the data subject? (Y/N)
Data correct	Was the data that the organization provided considered correct and complete by the data subject? (Y/N) Please note that this can still be true if no data is shared, when the data subject assumes it is correct that no personal data is found.
Answered	Did the organization answer all questions that are included in the request, besides a request for a copy of the data?
Specific	Did the organization answer questions specifically or generally?

Table 4. *Response variables and their corresponding criteria.*

The variables in this table all provide us with information on the response quality of the access request. Most of these also used for this goal in earlier research, with others were added during the data collection process when differences in responses were observed that could not be captured in available variables. In capturing these five variables, results are expected to provide a detailed understanding of organizations responses that goes beyond just the compliance of an organizations SAR responses. Although not all response measures

might influence the results of statistical tests in the same way, the details that each of the five variables brings is very relevant for the visualization and qualitative analysis of results.

For numerical analysis of the relationships that are stated in the first four subquestions, the dataset should be supplemented with factors detailing the response, judicial, organizational and subject factors that influence the data controller's response. The factors are represented by the variables that are described in Table 5.

Factor	Criterium
	Judicial factors
GDPR	Was the GDPR enforceable during the collection of the data? (Yes/No)
	Organizational factors
Age	For how long has the responding organization existed in its current form? (In years)
Size	How big is the responding organization?
	(Classification in small, medium and big organizations, following criteria from the Dutch Central Bureau of Statistics, CBS (2008)
Location	Where does the organization provides its services/goods?
	(Netherlands, European Union, International, International of Dutch origin)
Sector	What is the core sector of the organization's business relevant to this access request?
	Following the Eurostat SBI classification for sectors. (CBS, 2008)
	Subject factors
Relationship	What is the relationship between the data subject and controller?
	(Employee, ex-employee, customer, partner)
	Request factors
Intermediary	Was the request sent on behalf of the data subject by an intermediary party? (Yes/No)

Table 5. *Response variables and their corresponding criteria.*

In three of the factor categories, one factor is found which explains the influence of the entire factor group of a research question. The GDPR factor, which provides a distinction between requests that were answered while either the GDPR or DPD was in effect, also encompasses the effects of potential repercussions and organizations estimated risk in suffering these repercussions. Another reason for having just one factor in a category is because of research limitations. Only one request factor is included, because requests are deliberately executed in the same way to facilitate comparison of results with respect to other factors. In this way, the intermediary factor can be interpreted as a control factor. A larger research set-up might include further research into this factor. The subject factor group is also expected to consist of only one factor, the relationship that a data subject has with the data controller. Other factors might be of importance in this group but could not be tested in this research. Contemplations such as these are discussed in section 5, Discussions.

The effect of the organizational variable group is not expected to be solely explained by a single variable. Some or the organizational variables that are included in this research are chosen because earlier research has already shown them to be of influence in organization's responses. Examining organizations responses per sector for example, has already shown interesting results. However, since earlier research did not focus on examining other differences that might influence SAR responses in such detail, new factors are also included. These new variables are the included based on an expectation to provide interesting results, these factors are thus based on operationalized hypothesis.

3.3.2 Hypothesis

In choosing to include the factors of Table 5 in the research, their inclusion was expected to help explain at least one of the research's subquestions. Table 6 explicitly states the expected effect that each of these factors will have on the quality of an organization's response to an access request. The table also includes a motivation for each of these expected effects. When available, sources are given that substantiate this motivation.

The variables in Table 5 are selected on the expectation that these have an interesting effect on an organization's response to access requests. Table 6 lists the expected effect that each variable has on organizations' responses and why.

Factor	Expectation	Motivation
		Judicial factors
Regulation	Modern is positive	The GDPR presents stricter regulation and is expected to be more vigorously enforced. (Faradina, 2017; Ausloos & Dewitte, 2018; Norris et al., 2017)
		Company factors
Age	Negative	Older organizations operate older systems and have more data saved, which makes data collection for SAR requests harder. (Murphy, 2018)
Size	Positive	Larger organizations have more resources to correctly implement regulation and started preparing earlier. (Axinte, Petricã, & Bacivarov, 2018; "How the GDPR impacts and suffocates small and medium businesses," n.d.)
Location	Closer is positive	International organizations are harder to communicate with than local organizations. Organizations based outside the EU are less likely to follow EU regulation. (Faradina, 2017)
Sector	Competitive is positive	Specific relationships are more important to organizations in a competitive market. (EY, 2015; Sehested, 2018)
	technology oriented is positive	Sectors with more modern systems can more easily comply with access request.
		Subject factors
Relationship	Negative for broken relationships	Organizations have a higher incentive to respond to customers and employees than ex-customers and ex-employees, as these relationships are more valuable to them.
		Request factors
Intermediary	No effect	This factor should not affect results, because of measures taken in section 3.6

Table 6 Expected influence of included factors on organizations' responses to SARs

Combining the expected effects in Table 5 enables us to transform the first research subquestions into the following testable hypothesis:

H1: *The introduction of the GDPR has led to better replies from Subject Access Requests (SARs) compared to replies under the earlier DPD*

H2: *Younger, smaller organizations in more competitive and technology-oriented sectors will respond better to SARs.*

H2: *Organizations will respond better to SARs when it concerns a data subject with a more recent relationship with the organization.*

H4*: Dutch organizations will respond better to SARS made by Dutch citizens, EU organizations, will respond better than organizations from outside the EU. Responses by Dutch internationals will be classified between the Dutch and International organizations on average*

H5: *The intermediary variable should have no effect on organizations response quality.*

A "better" reply to access requests in these hypotheses means that organizations are more likely to respond and more likely to be compliant and polite in these responses.

3.2.3 Quantitative methods

As reported in the first section of this chapter, a mix of quantitative and qualitative methods will be used to answer the research questions that are formulated at the end of chapter two. In combining these methods, relationships between variables can be reported with precise effect sizes and ranges of uncertainty while being supplemented by more detailed descriptions of subjective experiences and examples of good and bad practices. Another strength of this mixed methods approach was apparent in dealing with challenges that were encountered during the data collection phase of this research. The novel data collection method of exercising Subject Access Requests (SARs) on behalf of volunteers required a larger time investment than was initially expected. Researchers were thus not able to finish collecting and coding data resulting from both data collection waves. The unfinished data collection mean that response rates and the quality of answers contained in the second data wave are not a correct representation of reality. Quantitative analysis is therefore only based on results from a combination of the first GDPR data collection wave and the dataset by used by Asghari et al. (2017) under the previous Dutch data privacy regulation, the wbp.

This quantitative analysis starts with a visual analysis of the variables that are coded in the dataset. Trough visualization, interesting differences and interactions between variables can be found that not previously known. Visualization will be performed using the pandas, matplotlib and seaborn packages in python scripts executed through Jupyter notebooks. Differences in variables that are observed in this first step or expected to be present in the data following the hypothesis detailed in the previous subsection, will be statistically tested for significance. From the statsmodel and scipy python packages, Chi^2, Wilcoxon, Mannwhitneyu and t-tests are used (Lalla, 2017).

The results of these first steps already allow us to make some educated assumptions on how organizations respond to subject access requests and how the introduction of the GDPR influences these responses. However, it does not yet allow for confidential answers to the research questions posed in chapter 2. The division of results in 5 different variables provides great details in organization SAR responses, but also complicates the further analysis and interpretation of effects that variables have on these results. As response variables, all 5 variables should describe a different dimension in the measurement of the same single underlying variable: how good a certain reply is. This assumption is tested by means of a Principal Component Analysis (PCA), in which an algorithm tries to find the best way to describe results by using fewer variables.

Through an interpreted PCA, the five dimensions that measure organizations responses can indeed be reduced to two binomial variables. The two variables are further combined into one ordinal value. Since the PCA method is performed in a very exploratory nature, the specific process of is best explained by example. This explanation can be found in annex A.5.2.

The two dimensions resulting from the interpreted PCA, provide a good starting point for the last quantitative analysis step: the interpretation of two binomial logistical (logit) regression models. By including all relevant factors in this model, clear conclusions can be drawn on the effect that these factors have on the specificity and compliance of companies' SAR responses. An added benefit to the use of these logit models is the easy interpretation of the rate with which each factor influences these rates.

3.2.3 Qualitative methods

Although unfinished, the second wave of data collection has led to interesting new insights into both the process and responses following from Subject Access Requests (SARs). The qualitative analysis of section 4.3 is therefore based on a combined set of data from the wbp data and both GDPR collection phases. This analysis has two goals. Because it is based on a larger dataset, the first goal of the qualitative analysis is to test if the relationships that were found in the quantitative analysis are also seen in the additional data. By revisiting these relationships in light of new data, the quantitative analysis will also serve in providing examples of practices that illustrate or contradict the quantitative findings. Focused analysis of these cases can provide the background information that is needed to create informed policy advice.

Often, the strength of quantitative methods is not found in the testing of hypothesis, but rather in the generation of new hypothesis. This is the second goal of qualitative analysis performed in this research. In searching for possible factors influencing certain behavior, the discourses of denial framework is applied. This framework is designed and implemented by Norris et al. in their influential 2017 publication. The framework describes six types of restrictive practices encountered during the data collection that formed the basis of their research. Categorizing practices encountered during the data collection under the GDPR in the same manner provides a proven manner of organizing SAR responses in a qualitative manner which also facilitates the comparison of these responses with experiences under previous regulation.

Discourse	Explanation
Out of sight	The data subject is unable to contact the data controller
Out of court	The data controller does not recognize the subject's right
Out of time	Data controllers employ delaying tactics aimed at discouraging data access attempts, or even wait by sending a reply until the data is erased (automatically)
Out of order	Administrative deficiencies make a request or response unavailable or excessively burdensome
Out of tune	Data controllers restrict access requests by refusing to deviate from internal procedures, even if these do not deliver the insight required by law
Out of mind	The data subject is treated as mad or having bad intentions by submitting access requests,

Table 7. Discourses of denial as used in Norris et al. (2017).

3.3 Scope

The scope of this research is dependent on the volunteers and the organizations that these volunteers allow to be included in the research. Given the personal nature of examined data and the resources that are available to the researcher, most of these volunteers will be close to the researcher in a personal nature. The scope of this research is therefore set to organizations that are expected to possess personal information of data subjects close to the researcher, meaning subjects that are all Dutch, most often in their 20s and predominantly highly educated. the scope will be limited to organizations that deal with (these) Dutch citizens.

Since data subjects share their personal information with hundreds of organizations, the volunteer sample is still expected to yield organizations in all shapes and sizes. To not overload a volunteer, only around ten of these are selected for research. The total dataset that is formed by response to access requests should provide enough insight to both answer the research questions and reflect on the impact of the GDPR from a societal point of view. The group organization sample is therefore chosen on the of the following criteria:

- The final dataset should include responses of organizations with as much different characteristics as possible, to analyze the effect of this characteristics.

- Organizations that are expected to possess more personal data or personal data of a more personal nature are regarded as more interesting for research into access requests, as non-compliance of these organizations provides data subjects with a higher privacy risk.
- Organizations that are expected to have personal data and perform a (semi-)public service or task or are otherwise (semi-)mandatory to be used in daily life are also seen as more interesting as these organizations are not expected to be scrutinized by most. Examples include utility providers, notaries and pension funds, which ordinary citizen are expected to share information with at some point in their life.
- Organizations that the volunteers see as more interesting are better research subjects, as these increase the investment of volunteers in the research.
- Very small organizations (e.g. independent entrepreneurs) might be excessively burdened by an access request. These should only be included if they are expected to provide meaningful insight that might not be found without this organization.

Of course, it is might be hard to select organizations that fit in all of these criteria. In the healthcare sector for example, many of the more interesting data controllers are small organizations such as GP's, dentists and other professionals. The criteria will therefore be used as a guideline, rather than as a rigid rule. Still, following these guidelines ensures that research will encompass organizations that have both scientific and societal relevance, while foregoing easier "low hanging fruits" such as popular international tech organizations which are already widely discussed in media pieces on the GDPR.

The scope of this research is thus defined as a group of 100 to 300 organizations that are expected to process data of Dutch citizens, are not excessively burdened by access requests, and are novel as well as interesting as a subject from both a research and a societal point of view.

3.4 Bias and verification

By performing this thesis research in the proposed form, the researcher is aware of a set of biases that accompany the data collection method. The main bias that is expected in this research is an unnatural response from organizations. When organizations know that their response will be compared with responses from other organizations, some may be fearful of the consequences that a non-perfect reply will bring to the respondent or his/her organization. This bias is inherent to research into data privacy rights in practice and is also reported by earlier researchers. It therefore does not necessarily affect the relationships of factors found within or between datasets but might affect the generalization of results to the real world. (Norris et al., 2017)

The risk of the unnatural response bias is higher for the GDPR dataset, because the novel data collection method through acting on behalf of volunteers makes scrutiny into its purpose more likely. Since it is deemed unethical to outright lie to organizations about the research purpose with which their responses will be processed, but counterproductive to state the purpose outright, the following steps are designed to deal with the bias as much as possible:

In the original Subject Access Request, no obvious links are presented to Delft University of Technology or further research purposes. The communication consent form, in which the volunteer states that the researcher may act on his behalf to perform access requests, is a personal permission to this researcher. The form states that the researcher may also receive and safe the volunteer's personal data that is send in reply to the request but may only process this data when further permission is granted. The permission to use this data for research purposes is stated in the separate research consent statement, which is not attached in the initial request.

Without direct statements, the underlying research purposes of access requests are still easy for data controllers to observe indirectly. This can either be done by investigating the return address of the request

(which is the TU Delft faculty postal address), reading about earlier research on the datarights.me homepage or looking up the researcher online. These routes are consciously left open for data controllers that are correctly investigating the request's purpose as it would be unethical to further cover the data processing purposes from a processor's due diligence. When data controllers do take this step and see the research purpose, it is very likely that they share their findings with the researcher, either for transparency purposes in data or as a reason to further investigate the volunteer's consent to this practice.

When such a situation occurs, the data controller's findings will be answered by providing the volunteer's research consent statement and an explanation as to why this was not attached to the initial request. In this answer, it is also explicitly stated that no specific organization or respondents will be identified in this research's results. Special emphasis is placed on a request to process the request as similar as possible to a regular request.

It is assumed that employees who take the extra effort to research the nature of the request and are assured that mistakes will not be traced back to them or their organization will have no problem to process the request the normal way when asked. However, to check if the novel data collection methods did indeed not influence results, a verification dataset was created. This verification set consists of access requests made personally on behalf of the researcher to organizations included in the research.

Another potential bias lies in the assumption that an organization will always respond in the same way when all variables included in the conceptual model of chapter 2 are constant. The differences that will be experienced in this case should be captured by the noise variable that is added in the regression models of chapter 4 (as "Intercept").

4 Results

As detailed in the previous chapter, the results of this chapter are based on a combination of an existing wbp dataset and one that was collected using a novel intermediary method. This novel collection method was applied in multiple phases, with a first wave of 8 participating volunteers and a second wave consisting of 25 volunteers. Since the data collection phases of both volunteer groups required a larger time investment than was initially expected, the second wave could not be finished before the writing of this report. The unfinished data collection means that response rates and the quality of answers contained in the second dataset are not a correct representation of reality. Quantitative analysis is therefore only based on results from a combination of the first GDPR data collection wave and the wbp dataset. This quantitative data concerning organizations' responses to access requests under the GDPR and wbp is described in section 4.1 and analyzed in section 4.2.

Although unfinished, the second wave of data collection has led to interesting new insights into both the process and responses following from Subject Access Requests (SARs). The qualitative analysis of section 4.3 is therefore based on a combined set of data from the wbp data and both GDPR collection phases. After presenting the findings of the qualitative and quantitative research, section 4.4 will present the verification and validation steps that have been performed with the help of two further datasets. All findings will be put into their broader societal context in chapter 5, before overall conclusions are presented in chapter 6.

4.1 Quantitative analysis

Through the use of descriptive statistics and construction of logistical models, many of the hypothesis that were formulated earlier could be tested quantitatively. The results of the quantitative analysis are summarized and compared to the relationships that were expected in chapter two in Table 8. These and other findings will be presented in more detail in the following subsections. Section 4.2 will test the same relationships in a quantitative analysis.

Factor	Expected relation	Relation observed
	Judicial factors	
Regulation	Modern is positive	The GDPR does indeed have a big positive effect on both compliance of organizations responses to SARs. However, responses also seem less specific because of the judicial change.
	Company factors	
Age	Negative	Could not be tested in quantitative analysis
Size	Positive	Negative, small organizations are both more compliant and specific in their responses to
Location	Closer is positive	Dutch organizations seem to produce more compliant responses. The effect is mostly because of Dutch organizations are smaller on average than international organizations tough.
Sector	Competitive is positive	No clear effect on the organization's compliance is found. Organizations in sectors such as ICT and other services seem less likely to respond in a specific manner, but there's no conclusive effect in other sectors.
	technology oriented is positive	
	Subject factors	
Relationship	Negative for broken relationships	Could not be tested in quantitative analysis

		Request factors
Intermediary	No effect	Could not be tested quantitative analysis Addressed in section 4.5

Table 8. *Expected and observed relationships of variables on responses to SARs*

4.1.1 Descriptive statistics

An important first step in the data analysis following to the findings presented in Table 8 is a description and visual assessment of the distribution and relationship of different variables. When such relationships seem plausible and present, statistical testing is used to check for significance. Unless reported otherwise, a p value of .05 is chosen as a threshold of statistical significance. This subsection will present the most interesting and important relationships that were found during data analysis. A full presentation of steps undertaken during data analysis, including visualizations of all variables and most interactions can be found in annex A.5.

As previously written, quantitative analysis is undertaken on a combined dataset. This first part of this dataset describes responses of 99 organizations operating under GDPR jurisdiction and was collected in the first wave of the data collection that is discussed in the previous chapters. The second part consist SAR responses from 93 organizations under the wbp (Asghari et al., 2017). Figure 2 features the number of organizations in these datasets per sector. Each sector group consists of enough organizations in the total dataset for analysis purposes. However, some categories in this larger set are dominated by one of the underlying sets.

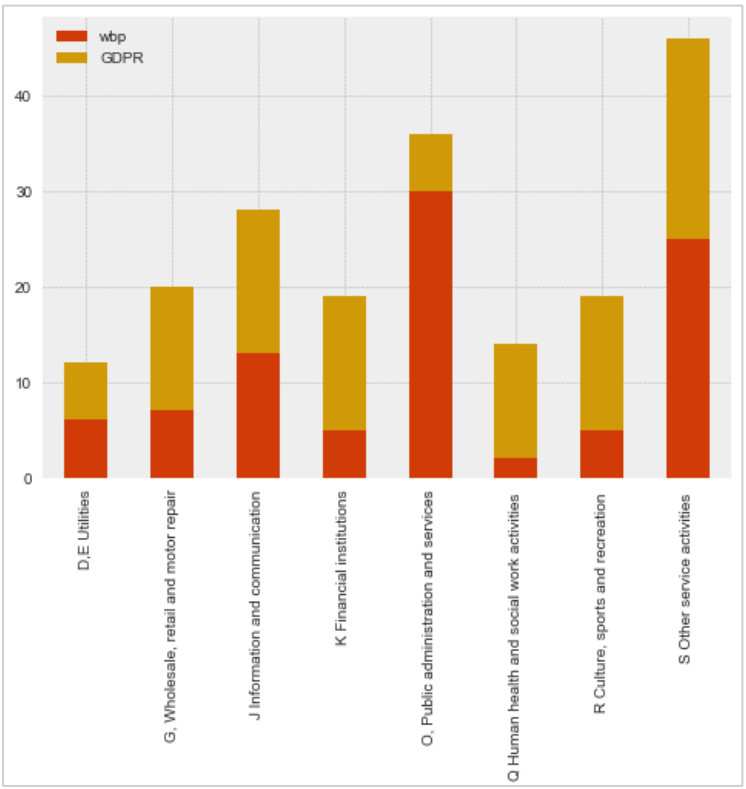


Figure 2. *Visualization of the number of organizations included in the numerical analysis sector, per dataset. Sectors are coded using the CBS SBI codes. Small sectors with comparable organizations are combined into bigger sector groups, as detailed in annex A.5.*

The K, R and Q categories feature fewer than 6 organizations operating under the wbp regulation, which would seriously hinder numerical analysis of these sectors under the former regulation. The unevenness between the relative share of organizations from different dataset in sectors in is important when interpreting results based on sectoral differences.

In figure 9, the total number of organizations is categorized by size, as defined by the Dutch bureau of statistics (CBS, 2008). A big part of the dataset consists of large organizations (>250 employees), but this is not unexpected when sampling organizations out of suggestions made by volunteers. Although many smaller organizations exist in the Netherlands, their smaller customer base makes them less likely to be included in a sample. This pattern is observed in both underlying datasets, although the difference is more extreme in the wbp set, as seen in figure 10. This might make meaningful analysis of differences in responses between small and other organizations more difficult without controlling for the difference in jurisdiction.

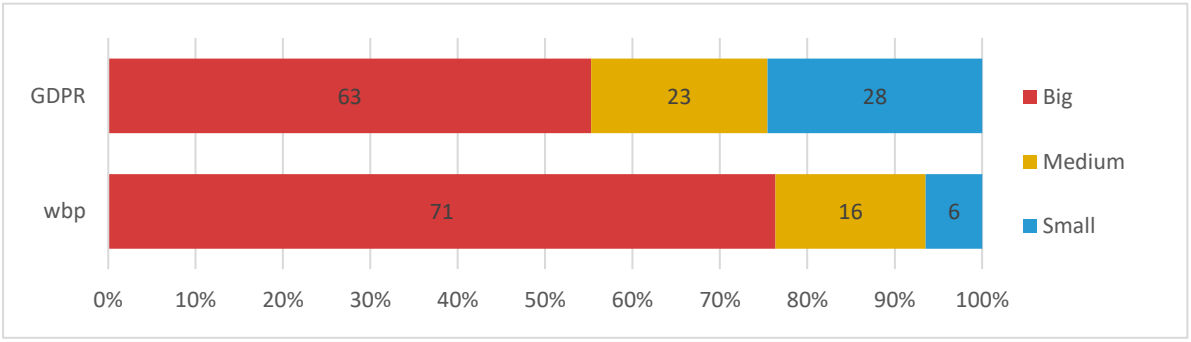


Figure 3. *Visualization of the size of organizations as percentage in both underlying datasets. The amount in each category is presented by the numbers within each bar.*

Differences in organizations' operating locations are also expected to be influential in their response to subject access requests. As can be seen in Figure 4, over 60% of the organizations in the dataset are only operating in the Netherlands. With a 20% share, the second biggest group in the combined dataset is Dutch organizations that operate internationally. The number of EU or other International organizations is much smaller with a presence of only 11 and 6% respectively. This distribution is probably a result of the sampling strategy, which is explained in section 3. When examining the differences in the location distribution per underlying dataset, the combined share of 80% Dutch and Dutch international organizations is overwhelmingly present in both. The low number of international organizations, especially the minor three big sampled in the wbp set, makes analysis of organizations inside and outside the European Union challenging.

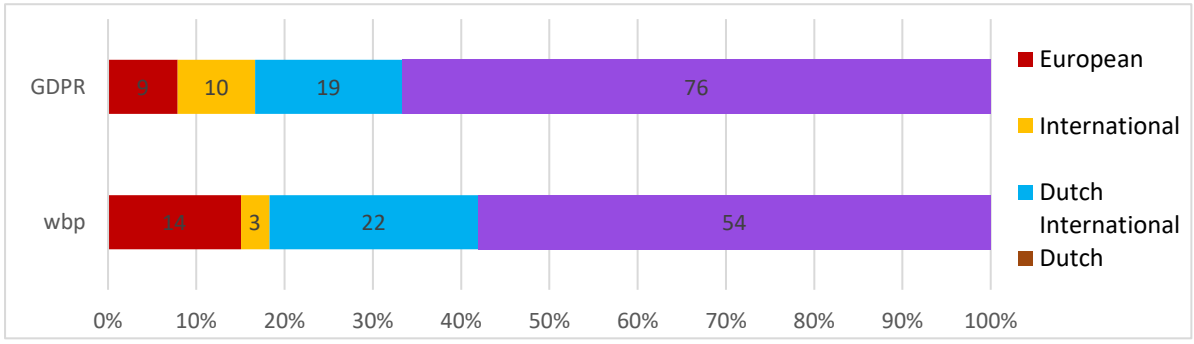


Figure 4. *Visualization of the operating locations of organizations as percentage in both underlying datasets. The amount in each category is presented by the numbers within each bar.*

The response of organizations to Subject Access Requests (SAR's) is coded in five variables. Figure 5 shows the percentage of organizations that fulfill the requirements represented by these variables divided among by the two underlying datasets, characterized by the regulation factor that differentiates these sets.

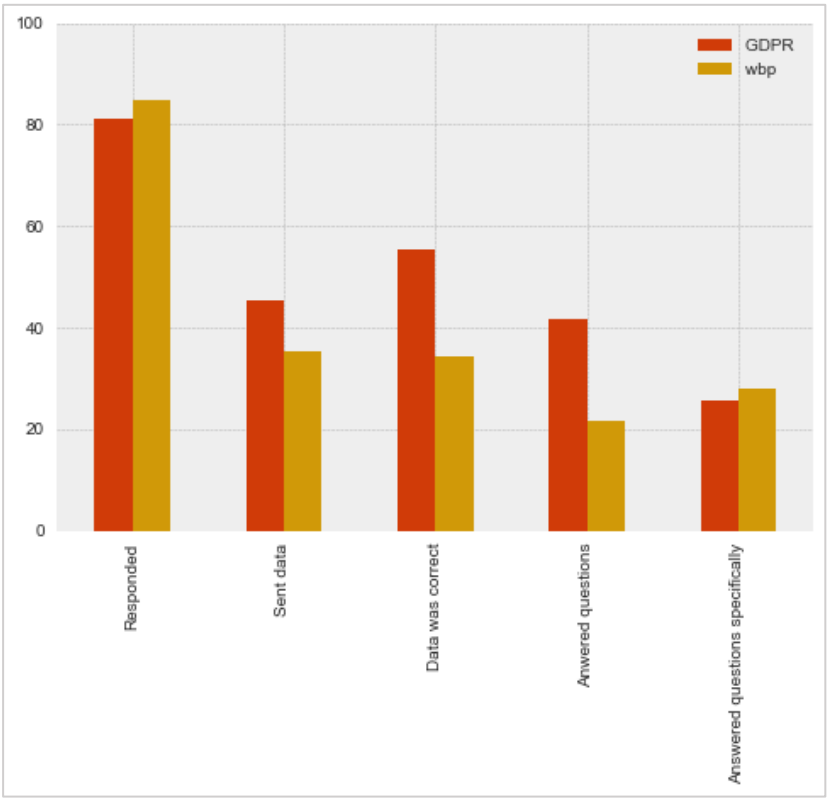


Figure 5. *Percentage of organizations that responded, sent data, sent the correct data, answered further questions and answered these questions specifically under the different regimes*

The average response rate of organizations replying to access requests is around 83%. The response rate is a bit higher in the dataset capturing responses under the wbp (81% vs 85%). This noticeable but statistically insignificant (p=0.24) difference is not expected but can be attributed to a difference in data collection between the two datasets, which is discussed in chapter 5. The second response variable presented in figure 5 presents the percentage of organizations that did respond to SARs by sending personal information in some form. The bar graph shows a clear difference between the 35,5% of organizations that did so under the wbp and the 45,5% under the GDPR. The p-value of 0.078 that follows from a one-sided, paired T-test between sent data rates of organizations operating under the different jurisdictions shows a 92% certainty that the rate is indeed higher for those operating under the GDPR.

Personal data that was sent by organizations is not always correct (e.g. it is incomplete or inaccurate), and organizations that do not supply personal data are not always incorrect (e.g. the organizations has already deleted the data). Data subjects are therefore also asked to judge if the information that was (not) sent by organizations was a complete and accurate representation of the personal information that the data controller was reasonably expected to possess. The difference of 21% in this data correctness rate is the biggest difference between responses by organizations operating under different regulation.

It is important to note that the size of this variable is thus based on assumptions by the data subject, although often informed by the researcher which can connect the data that was sent to responses of comparable organizations. Cases in which answers were assumed to be incorrect are discussed in the qualitative results section. Limitations in the interpretation of this and other results are discussed in more detail in the discussion section.

Obtaining a copy of one's personal data is not the only way in which organizations should comply in providing insight into subject's data use. Both the wbp and GDPR specify further information that organizations should supply when requested. The "Answered questions" variable measures how many organizations have responded to all of these questions. Even though organizations in the GDPR dataset were asked to answer more questions, a higher number (42%) of organizations answered all of these compared to those in the wbp set (22%). Out of the four response differences presented in figure 15, the difference in this variable is the most significant, statistically speaking, with a p value of 0.001.

Broadly speaking, the organizations' responses that answered all further questions asked can be divided into two categories. The first of these is are specific responses, in which the questions asked are answered uniquely regarding (only) the personal data that it has processes on the volunteer. The second category consists of general responses. This means that the response included general information about all data processing which is carried out by the controller (out of which the data subject has to infer responses relating to their situation). These two categories of responses are captured by the last variable displayed in Figure 5. Organizations acting under wbp regulation often responded in a more specific manner to the posed questions than those under the GDPR, although this relation is the least statistically significant of the response variables with a one-sided independent t-test reporting a p value of 0.36. Organizations in the wbp dataset are thus more likely to answer specifically, but incomplete, while organizations in the GDPR dataset are more likely to answer complete, but general. This issue will be analyzed further in the next section.

4.1.2 Reduced dimensions

The results that are presented in Figure 4, already allow us to make some educated assumptions on how organizations respond to subject access requests and how the introduction of the GDPR influences these responses. However, it does not yet allow for confidential answers to the research questions posed in chapter 2. These 5 variables allow a description organizations' responses to subject access requests in more detail. In other words, the variables present a different dimension in the measurement of a single underlying variable: how good a certain reply is. However, the division of results in 5 variables complicates both the analysis and interpretation of effects that variables have on these results, especially when examining differences in more granular variable, such as sectors. Through an interpreter PCA, the five dimensions that measure organizations responses have been reduced to two: compliance and specificity, which can both be described by binomial variables. Since compliance is seen as more important in answering the research questions of this thesis project in a societal perspective, the two variables can be combined into one ordinal value. This transformation is shown in Table 9. The distribution of the response score in percentages of organizations is seen in Figure 16. The process of interpreted PCA dimensional reduction is described in more detail in annex A.4.

For the interpretation of values following from these reduced dimensions, it should be noted that the specificity variable is partly based on the response rate of organizations. the compliance measure is also affected, since the response rate indirectly influences all other measures. The response value is known to be measured somewhat differently between the wbp and GDPR datasets because of a difference in data collection regimes. Because of this difference, scores of organizations operating under the GDPR can be lower than organizations

	specificity	
Compliance	High	Low
Yes	4, Excellent response	3, Adequate response
No	2, Partial response	1, Failed response

Table 9. *Two binominal dimensions are found in the results through interpreted PCA analysis. This table shows how the two dimensions are combined to create a single response quality variable*

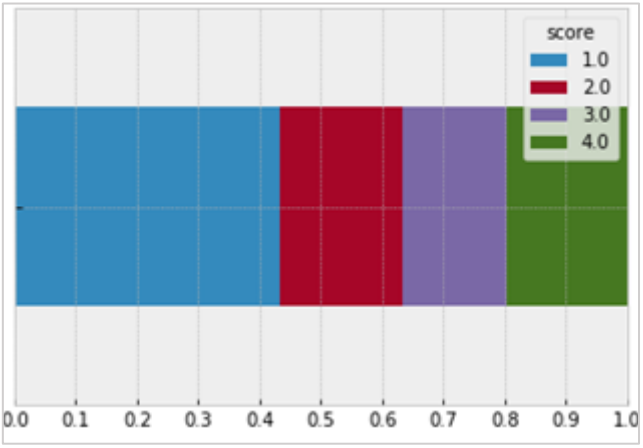


Figure 6. Visual representation of the SAR response quality of organizations as a share of the total number of organizations.

As Figure 6 shows, the "failed" group is the biggest in the dataset, with 41 percent of organizations being labeled both noncompliant and not polite in their handling of subject access requests. On aggregate, 35% of organizations is compliant, with 19% responding excellent. The dimensional reduction provides an easy way to visually investigate the effects that organizational and judicial variables have on an organization's response to subject access requests. These visual effects, when combined with statistical testing provides great insights into the effects that are hypothesized in chapter 3. However, the probable correlation of organizational factors and small number of observations for some values dictates caution in the interpretation of these direct effects.

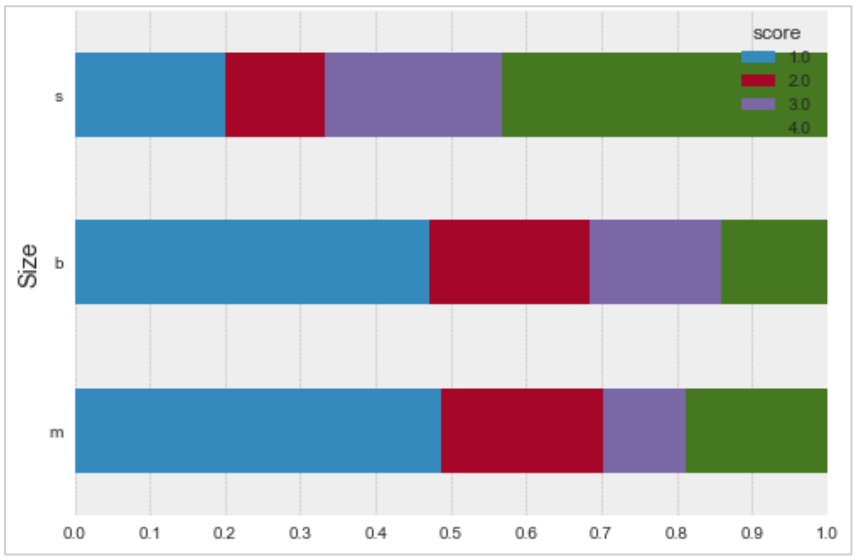


Figure 7 Visual representation of the share of the quality of SAR responses by organizations of different sizes.

Contrary to the hypothesis formulated in chapter 3, smaller organizations seem to respond superior to subject access requests, having double the compliance rate compared to bigger counterparts. Indeed, the difference between medium and big sized organizations is not statistically significant in the nominal response and binomial compliance variables. The group of small organizations does show statistically significant differences with both other groups.

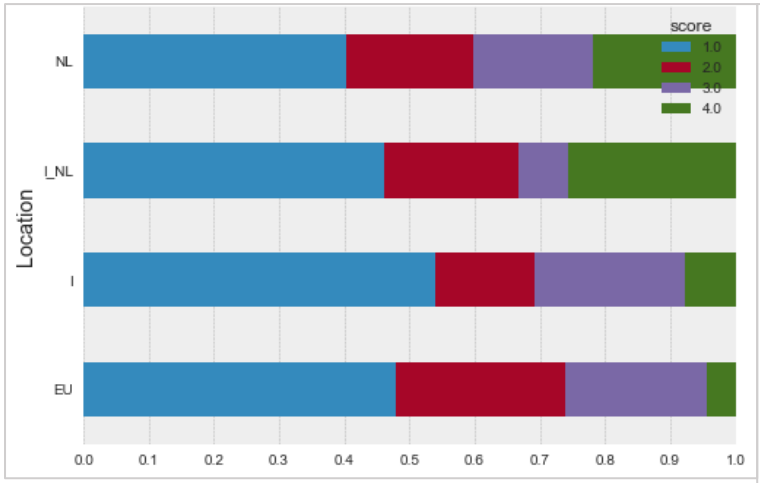


Figure 8. *Visualization of the share of SAR responses of different quality, grouped by location of operations.*

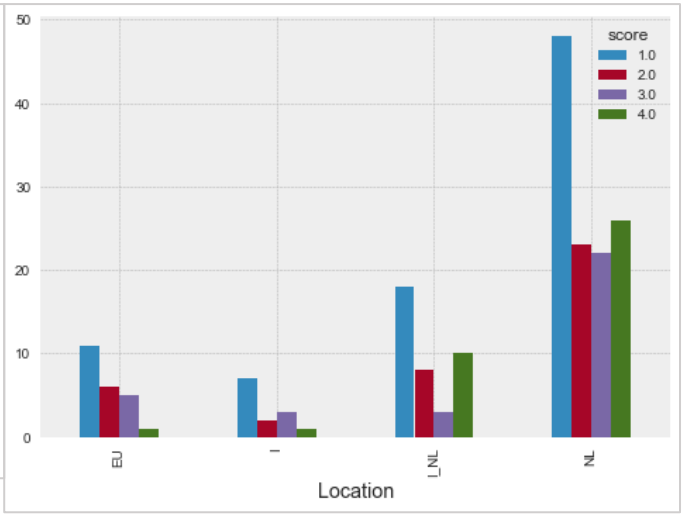


Figure 9. *Visualization of the number of SAR responses of different quality by organizations, grouped by location of operations*

Dutch organizations seem to respond better to access request than their European and international counterparts, following visualizations in Figure 8 & 9. International organizations originating in the Netherlands are not significantly more compliant in their responses than other international organizations, but certainly seem more specific in their compliant responses. Figure 8 suggest that International organizations actually outperform European organizations in compliance terms. This difference cannot be verified statistically. Figure 9 also visually shows that this comparison between these organizations in figure 8 is based on a small sample size.

Sectoral differences in compliance and overall response classification are also apparent and displayed in figure 10. Due to the uneven division of sectors per dataset displayed in figure 2, the results cannot be properly interpreted without controlling for judicial differences.

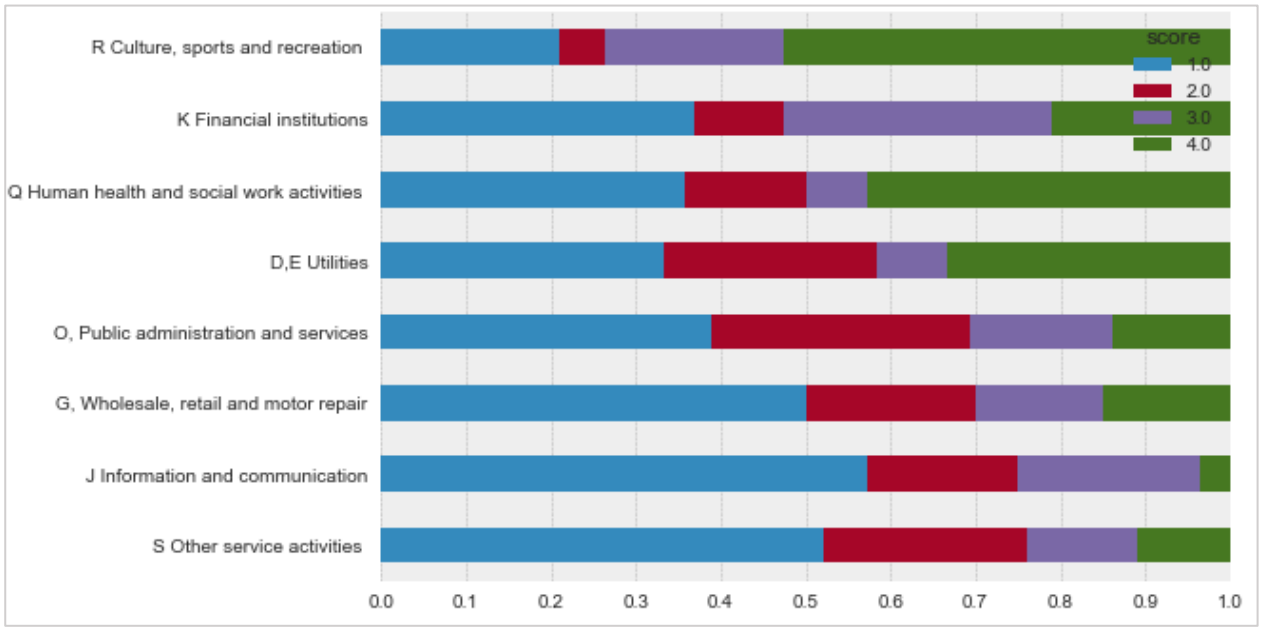


Figure 10. *Share of organizations of different response score, presented per sector as defined by the CBS (2008) SBI.*

4.1.2 Regression modelling

Judicial and organizational factors are shown to probably influence both the quality of organizational responses to SARs. In order to accurately describe the effect that all these factors, regression models are created. In these models, the compliance and specificity of responses are presented as independent binomial variables that are influenced by the dependent judicial and organizational factors. For both response quality dimensions, the most important measures following from the "best" model is presented for interpretation. This model is chosen on the basis of its AIC score, a measure which balances the relative explaining power of a model with the number of included variables. (Yamashita, Yamashita & Kamimura, 2006) More detailed results and other models that were created during the data analysis section of the thesis can be found in annex 5.

Table 10 shows the results of the best model for explaining organizations' SAR response compliance. Since only the small organization size and GDPR factor add significant explanation of organizations' SAR reply compliance, the other variables have been omitted. Since the logit regression method models the logistic effects of binomial factors, the coefficients of this model can be interpreted as odds ratios. These ratios are translated to effects sizes in Table 10. A unit increase in the GDPR variable, increases organizations compliance rates by 3.1 times, a growth of 210%. This GDPR variable constitutes to the difference of organizations operating under the wbp or GDPR. The results in Table 10 thus show that the introduction of the GDPR has increased the compliance rate by more than 3 times. The 95% uncertainty interval of the effect lies between an increase of 64%, and one of 500%.

	Coefficient	**P > |Z|**	**Effect size**	**95% certainty effect**	
GDPR	1,1436	0,001	3,138	1,641	6,001
S	1,1976	0,006	3,312	1,401	7,828
Intercept	-1,394	0,000	0,248	0,149	0,412

Table 10*. The estimated coefficient, probability value, effect size and 95% uncertainty interval of this effect of the best logit model explaining organization's SAR response compliance*

In similar fashion, a unit increase in the S variable, increases organizations compliance rates by an increase of 230%. The 95% confidence interval of this estimate shows that small organizations are between 1.4 and 7.8 times more compliant than organizations of a different size in this model. A high standard error makes the precise size of the GDPR and small organization's effect unclear. However, the model clearly shows that both variables have a big positive impact on organizations SAR response compliance. The final variable in Table 10 is the intercept. This variable represents the expected value of the model if all other factors are 0. It shows that the mean response compliance score of organizations not operating under the GDPR or having more than 25 employees (and thus not being a small organization) score poorly in the response compliance measure. This further affirms the conclusion that the GDPR and small size have a large positive influence on organizations SAR compliance.

A similar logit-model was created for the specificity of responses. The model's most important results are shown in Table 11. It shows that the introduction of the GDPR has decreased the likelihood of specific replies by 63% on average, with a 95% confidence interval showing a decrease between 21% and 82% in the specify measure. Smaller organizations are most likely expected to respond more specifically to SAR's than the model's average organization, but the wide confidence interval makes this interpretation somewhat uncertain. This 95% interval shows the effect to be between a decrease by 2% and an increase of 500%. With a P value of .055, the effect is clearly less significant than the compliance effect of smaller organizations.

	Coefficient	P > |Z|	Effect size	95% certainty effect	
s	0,887	0,055	2,427	0,981	6,005
GDPR	-1,013	0,005	0,363	0,178	0,740
K	-0,994	0,097	0,370	0,114	1,197
G	-0,961	0,097	0,382	0,123	1,189
S	-1,106	0,018	0,331	0,132	0,827
O	-0,879	0,090	0,415	0,151	1,145
J	-1,647	0,005	0,193	0,061	0,607
Intercept	0,787	0,061	2,197	0,965	4,998

Table 11. *The estimated coefficient, probability value, effect size and 95% effect uncertainty interval of the best logit model explaining organization's SAR response specificity*

Following the effect sizes reported in Table 11, organizations in the financial (K), retail (G), ICT (J), Public (O) and other services (S) sectors all seem to be less likely to respond to SARs in a detailed manner. However, most of the effect size of these sectors are coupled with a high p value and uncertainty effect including cases in which sectors in organizations are more likely to respond specifically. This makes the reported effect unclear. Organizations in the other services and ICT sectors can be concluded to be more specific in their reactions, albeit with large differences in the 95% certainty rate.

Including location and (other) sector factors does not add much explaining power to either of the models. This is probably because the variables are both correlated with an organization's size. Models that include these factors without including the size are still not very good in predicting an organization's response score, however. The uneven distribution that was observed in these categories in the previous subsections is probably another reason for its uncertain effects.

The relationships between an organization's SAR response quality and the factors expected to influence this organizations that are found trough quantitative analysis are summarized and compared to their expected relationship in Table 7, at the beginning of the chapter. These effects will also be investigated through qualitative analysis in the following section.

4.3 Qualitative analysis

As previously discussed, data was collected on organization's responses to Subject Access Requests (SARs) under the GDPR in two waves. Combined with the wbp dataset, this qualitative analysis encompasses 458 requests. The organizations that were added to this number in the second data collection wave were specifically chosen to balance the distribution of GDPR and wbp request responses in certain sectors. The qualitative analysis is thus not only used to verify the quantitative results, but also to provide further insight into the process of requesting access to personal data processing and factors that influence this process. To this extend, the section will start by comparing data collection experiences with those detailed by other authors exercising Subject Access Requests (SARs).

4.3.1 Discourses of denial

Results of the regression analysis of the previous section already elude to differences in the way organizations respond to SARs aside from their compliance with appropriate regulation. During the data collection process, the tone of organizations was perceived to be the biggest improvement on descriptions on the same process under previous regulation. This is analyzed by exploring the hurdles that were experienced by researchers and volunteers during the collection of this data. These hurdles will be compared to the six "discourses of denial"

of Norris et al. (2017), a classification of the hardships that were experienced during the collection of similar data before the GDPR came into effect.

The first discourse of denial, *out of sight*, is experienced when data controllers could not be contacted, or did not respond to inquiries. Encounters described by Norris et al. (2017) and others where an organization never replied to access requests are still widespread (although somewhat less frequent). Other cases of this discourse, where correspondence with larger organizations meant dealing with lots different departments were also still experienced. Even when a request was sent to the address that was explicitly named in the privacy statement as the location to send subject access requests to, some big multinationals still replied that the request should be sent to another address and would thus not be processed. Although examples still exist, this discourse of denial is less frequent with the GDPR, because larger organizations are obliged to have a Data Privacy Officer (DPO). When requests were denied by the customer service department, a complaint to the organizations DPO often received a more facilitative answer.

The second discourse of denial, *out of court*, describes a situation where a data subject's right to access his/her personal information is not acknowledged or unlawfully terminated because of incorrect interpretation of the law. This can be an effective barrier for data controllers to stop inexperienced data subjects in exercising their rights. In our data collection, the discourse was often met because of the specific data collection techniques in requesting access on behalf of others (which is discussed separately). This barrier is most often caused by data controllers' unawareness or unexperienced with privacy regulation. When the *out of court* discourse was experienced, this was often because the organization was too strict on the sharing of personal data, sometimes citing parts of the GDPR and sharing their fear of possibly causing something that could be labeled a data breach. For larger organizations, the mandatory DPO seemed to help a lot with the correct interpretation of the law. Most smaller organizations seemed to really try to follow the law correctly, but sometimes interpreting it too strict. Through interaction with some of these smaller data controllers, it became clear that they had never seen an example of what a correct response to a SAR should looked like and just applied the law as narrow as possible in order to avoid fines. In these cases, the introduction of the GDPR might have paradoxically both decreased subject's privacy rights and made smaller controllers with good intentions liable for fines.

Another discourse of denial is the *out of time* claim. A handful of organizations responded to a request by (lawfully) asking for an extension of two months on top of the first month that organizations have to respond to the request. Most of these organizations never responded after the extended deadline. From the perspective of a data subject (or researcher), three months is a very long time to wait for an organization's response to an access request. The extension can be useful for organizations that need the time to gather lots of personal data from different places before providing full insight. Most of the organization that asked for this extension however, did not use the time productively. If a response was received after three months of waiting, it often contained another discourse of denial, designed to buy even more time. Others claimed to have "forgotten" the request or explained that the request was "lost" somewhere in their internal process between two departments. Just like most other discourses of denial, a fraudulent *out of time* claim plays to the information asymmetry between data subjects and controllers. Collaborate efforts and the help of experts during the process may help break this asymmetry and expose dishonest claims through comparison.

Some examples of the *out of order* discourse was also experienced, highlighting organizational mis-management of the request, thereby creating a bad response. This was often experienced by organizations that use some sort of specific tool or questionnaire to receive access requests. Usually, these tools streamlined access requests to a point where a response was received very quickly if the data subject was able to provide all necessary information correctly. These tools often did not account for other questions that the subject is also allowed to ask about the processing of his personal data, or deliberately limited fields where such questions could be entered.

Other examples of these *out of order* requests were found in paper forms that were returned by organizations that received an access request. The data subject was then supposed to fill in this form and send it back to the processor before the request could be processed. The form was often supposed to be filled with information that was already shared in the original request and thus completely irrelevant or asked the volunteer and researcher to share a lot of personal information without detailing why this was necessary. Requests to provide a motivation for the necessity of this data often received a response such as: "Otherwise I can not start the process from my side", "This information is a necessary field in our database" or "If I leave it blank, our data will be tainted". Although some examples do exist of organizations that correctly implemented tools or forms which did facilitate easier data insights, numerous other implementations only increased administrative burdens. The vast majority of organizations that imposed these administrative burdens were public bodies. Municipal governments in particular were found to be very keen on employing this tactic.

Forms and tools are most often used by medium and large organizations to streamline their SAR process. But where medium sized organizations often do respond to further, specific inquiries and an appeal to the DPO of larger organizations often helps to find answers, the biggest organizations may not respond. When questions are sent that explicitly state that the tool used for data insight did not or only partly address the questions asked, these organizations would often reply with another link to their tool, telling data subjects that this form is the only way to request data insight. This *out of tune* behavior is not often observed in this study. This might be because it is already part of lawsuits against some of the biggest tech organizations in the world.

The sixth discourse of denial that is detailed by Norris et al. (2017) is the *out of mind* response, in which the data controller responds that insight should obviously not been given following the request. Some smaller organizations did respond in such a way during the data collection, signaling that they did not understand what would motivate a data subject to request insight into his/her data. Others small organizations, mostly those in the healthcare sector, responded that they would rather use their limited time to help people that were in direct need of medical assistance, than responding to the request. Although such a response cannot be seen as very facilitative, it was often worded in an unaggressive manner and accompanied by an acknowledgement that the data subject was within his/her rights to request an answer. When volunteers decided to further proceed with the request, the response was often very detailed and polite. *Another out of mind* practice that was observed was certain data controllers' distrust for the incentive of the received access request. This was observed in legal professions especially, but also in other sectors when excessive proof of identification was deemed mandatory. This distrust can be partly explained as due diligence of an uncommon way to exercise access rights (on behalf of others) and will be more thoroughly analyzed in the following section.

Previous examples show that each of the discourses of denial that were described by researchers requesting insight into their personal data under previous regulation, still exist in the GDPR era. The introduction of the GDPR clearly did not solve all existing problems with access requests, which can still be a burdensome task to complete. However, it is important to note that the previously mentioned examples constitute a small group of the worst practices that were received after contacting around 350 organizations. No responses were found to be obviously made in bad faith, and only two organizations responded that they had never heard of the new regulation. This is in stark contrast with responses published in earlier work (featuring smaller sample sizes), which details organizations denying receiving requests despite proof of delivery, straight out refusing to invest the required work to respond correctly or choosing to erase data instead of providing insight (Aulos & Dewitt, 2018; Norris et al., 2017; Herrmann & Lindemann, 2016). Rather, existing challenges seem to exist mainly because of internal processes that are intended to improve/streamline the very process they frustrate. The tools are used by employees that are either unknown to the GDPR or are to zealous in their implementation of the law. Of course, some of these responses might still be tactics for those that really do not want to comply and state a different reason.

4.3.1 Further observations

Following the more general conclusion that followed form the discourse of denial framework, this subsection will focus on other trends that were observed during the data collection process but did not fit the framework. This is because the observations are based on the research's focus on factors explaining differing responses between organizations and the novel data collection method.

During this data collection, two main challenges were observed that had to be overcome before any organization was willing to start its research process into the processing of a volunteer's personal data. The first was a proof of method. After sending a first group of SARs on behalf of volunteers, it quickly became clear that this method was not as common as previously thought. Only a small group of organizations knew that such requests can indeed be exercised on someone else's behalf when accompanied by a written authority. Many organizations first response was that they were only obliged to share data with the data subject and should not share it with anybody else, thereby referencing the GDPR and Dutch DPA. In contrast to the CIO, the DDPA explained this topic on the website as follows: *"The right of access only concerns access to your own data. You therefore have no right to information about others."* (Translated from: Autoriteit Persoonsgegevens, 2018b). After numerous calls were made and an official complaint was filed against an organization that categorically refused to co-operate with the request, this quote was extended to the correct: *"The right of access only concerns access to your own data. You therefore have no right to information about others. Unless someone has given you, for example, written consent for that action"* (Translated from: Autoriteit Persoonsgegevens, 2018b). The second was only added on the DPA website after most of the data collection was finished. Most of the collection process therefore started with an uphill battle to proof that the collection method was indeed lawful.

This legal battle did help to provide further insight into a second big challenge that often accompanied the first: proof of identity. Since data controllers should only share personal data with the data subject which the data concerns, it falls upon these controllers to ensure that the request is indeed made by (or on behalf of) this data subject. As is typical for SAR's, all access requests that were made for this research were accompanied by a safe copy of the identity card or passport of both the data subject and researcher. DPA's suggest that this is the most extreme measure to check the identity and that most organizations should op for less privacy sensitive method. This suggestion is opposed to the proof in practice though, with data controllers seldom accepting the copy of an identification document as proof of identity. This follows from GDPR article 64, which states that a data controller is obliged to use "all reasonable measures to verify the identity of a data subject who requests access" (GDPR, art. 64)

Most data controllers do not accept a copy of an identity card to be within all their reasonable measures to verify the identity of a data subject. This is because identity cards are designed to proof a person's identity in real life and its security features are often constructed in a way that prevent them from being copied. This makes scanned identity cards very easy to counterfeit. The copy of an Identity card thus often meant the start rather than the conclusion of negotiations between the researcher, volunteer and organizations over further proof that was acceptable for all parties. Many organizations accepted a confirmation from the data subject from an (email-)address or phone number known by the controller as a reasonable measure or found sending the response via registered mail as justification. Other organizations, however, were very strict in their interpretation of the tools that were within their reasonable means. In the reasoning of one of these data controllers, all attempts to contact the data subject could be seen as unlawful use of the subject's personal data if the subject did not authorize the initial request and all means that were available to the data subject' could not be accepted as proof since these could be falsified. Many organizations concluded that the only way to identify data subjects with certainty was for them to proof it in person. This conclusion was reached by which often never had any personal contact with the subject before. For some reason, the identity verification in organizations is often far stricter when it comes to requesting data than when it comes to submitting it.

Earlier publications also describe challenges related to identifications of data subjects in SARs and the *visibility paradox* that comes with this requirement: data subjects are often required to reveal personal information to the controller that was not earlier in the controller's possession to prove their identity. Citizens are thus forced to give up some of their privacy in order to exercise control on their privacy. With the introduction of the GDPR, this paradox seems to have increased to greater levels with many organizations compelling volunteers to prove their identity in person, even when the data controller is housed in the other side of the country. This practice is a clear result of the data privacy hype that surrounded the introduction of the GDPR. Organizations' perception of repercussions that will follow when information is wrongly disclosed are so high, that it will not even be shared with the subject it concerns. Data controllers fail to see that by opposing the publication of this personal information and its surrounding practices makes them in fact in violation of the very law that is used to their practices.

One interesting observation concerning the identification challenge is that organizations often did not oblige researchers to proof these two points when there was no personal data being processed of the volunteer in question. Sometimes organizations even disclosed this fact during a discussion on one of these two main questions: "By the way, I don't think it is necessary to discuss the subject any further. I checked and [volunteer] is not known in our database." Organizations were thus almost never willing to admit the existence of personal data without proof of method and identity, but keen to disclose that the volunteer was not known to them to get out of the discussion. This way, misuse of intermediary consent can thus still be used to create an overview of the organizations that do have someone's personal data by means of elimination.

Another important goal of the GDPR is to improve the communication of SARs. Literal quotes from the regulation show a clear preference, if not obligation for data controllers to be able to receive and respond to RARs in digital form:
The controller should also provide means for requests to be made electronically, especially where personal data are processed by electronic means. (Recital 59)
The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. (Art. 12:1)
Where the data subject makes the request by electronic form means, the information shall be provided by electronic means where possible, unless otherwise requested by the data subject. (Art. 12:3)
Where the data subject makes the request by electronic means, and unless otherwise requested by the data subject, the information shall be provided in a commonly used electronic form. (Art. 15:3)

This obligation seems to not be that important to data controllers in practice. The contradiction is exemplified by a public services institution that provides an online tool to create SARs but requests subjects to print the final form and send it via mail. Around one in ten organizations could only be reached via letter. Surprisingly, almost all of these organizations are among the biggest sampled in the research. The group mostly included large telecommunications providers, postal services and many public services which are almost guaranteed to process data via electronic means and have the resources to receive requests "by electronic means". Most other organizations accept electronic requests, with around 70% accepting this via email and the remaining 20% asking to receive the request via some online form (which was often limiting the request, as discussed in the previous subsection).

In all requests, organizations were specifically asked data controllers to send the information in "*secure form, in common, structured and machine-readable format, but in any case, by ema*il". Most replies containing personal information were indeed received as secured attachments to emails, with keys send in a separate email or transmitted via text to the researcher's or data subjects' mobile phone. Another relatively common practice, which was often used by organizations in the healthcare or insurance sector, was to provide SAR responses as downloads from a secure online environment. These environments often featured two-factor authorization as well as a limited to the number of downloads or time before the link was no longer valid.

Only a small subset of these replies contained a machine-readable file format though, with personal data often shared as pdf files. Most of these files contained the personal formatted in tables, which could have also been send as csv files. Some organizations shared personal data via pdf's containing screenshots of pages in database programs. When asked about this practice, one data controller responded that there was no way to retrieve the data in any other way known to him/her. Although no reference was made to this right in requesting access, the results paint a bad picture for subjects right to data portability, which in GDPR article 20 states that data which was originally provided to controllers should be received in a machine-readable format.

Obtaining the information in a machine-readable format was out of the question for organizations that insisted in sending their response via mail. This practice was often higher because responses were sometimes sent directly to volunteers via registered letters. Still, responding via letter was obviously desired over electronic communication means by organizations in the public sector. Multiple data controllers in this sector were of the opinion that letters were the safest or indeed only safe way to transmit personal data since one could never be sure where emails would end up.

4.3.2 Hypothesis revisited

In light of the qualitative analysis above, some hypotheses have been strengthened, while others have been weakened. This subsection will expand on the observations stated above and relate observations to the hypothesis.

Concerning the introduction of the GDPR, its effect is also clearly observed in a qualitative analysis. In general, this effect is positive. Few data controllers seem unaware of the law and the general idea of SAR's, which provides major improvements to the tone of responses. None of the analyzed organizations responded in bad faith and a decrease in the occurrence of almost all discourses of denials is observed. In response to the GDPR, many organizations seem to have created internal processes to manage SAR's. These processes may also present a weakness however, since some confine data subjects to ask for only some of the insight that is rightfully theirs. Personal responses to SAR's are often found to be more facilitative.

The awareness that the GDPR raises on the topic of data protection also has a darker side. Many organizations have become too anxious to share personal data with even the subject of this data, for fear of sharing too much by mistake. This anxiety leads to excessive barriers for subjects to effectively practice privacy rights, especially in having their identity checked by al "reasonable" means available. Data controllers seem subjective in their awareness of GDPR, often disregarding obligations to receive requests in digital form and respond in machine readable formats.

Qualitative analysis sheds further light on the impact of organizational factors on SAR responses. Small organizations are seen as more facilitative in handling these requests, as these are often handled in a more personal manner. Larger organizations seem to have a contrasting tendency to streamline SAR responses as much as possible, leading to less facilitative and complete answers. Having a DPO often helped improve the data controllers' responses when earlier responses were not satisfactory.

Differences in the specificity of organizations that were observed in the numerical analysis are also seen in the qualitative analysis, although it is hard to say how much of these qualitative results are also caused by organizations' size. Organizations in the ICT sector are more likely to use (predominantly non-restricting) tools for SARs, thus supplying more general results. Distinctly different practices are also observed in the healthcare and public sector. The public sector shows a particular distaste for electronic methods which would facilitate the practice of data subject's privacy rights. Organizations in the Dutch healthcare sector are found to stricter and sometimes conflicting regulation regarding data subject's right of access to processing information.

No new relevant observations have been added through qualitative analysis regarding the effect of organization's location on SAR responses. Effects of subject factors are also not observed in the qualitative analysis because of a small number of responses to volunteers with different relationships with controllers than that of a customer/supplier role.

4.4 Validation and verification

The previous sections have presented interesting insights concerning responses to Subject Access Requests (SARs) and the factors that influence these responses. Before these findings can be put into a larger context or used to conclusively answer research questions, it is necessary to check the verify and validify the findings. In this thesis, validation and verification will be performed in the context of computer simulation. (Balci, 1994) In this context, verification is performed to confirm that the analysis is indeed correctly implemented. Validation is performed to check whether the results of the analysis accurately describe corresponding values in the real world.

On face value, both the quantitative and qualitative analysis produce verified results. Both are based entirely on the collected dataset and are performed using proven methods. The regression analysis is verified through many rounds of bug fixing and expert advice and produces results that, while interesting, are not improbable following the conceptual model of section 2. Furthermore, limitations to the findings following from the model are consciously added to the results in the form of p scores or certainty intervals. These numerical results are further validated through results from the quantitative analysis, which in turn are also compared with findings from earlier research. Since this is the first research into the use of access request under the GDPR, findings could not be compared with research that was expected to produce entirely similar results. Findings are thus compared to results in research under earlier regulation, which can be found in Table 3. Results are found to be in a valid range given the earlier results and (model) expectations of changes because of the introduction of the GDPR following thorough literature research.

Verification of the results proves more challenging, since a part of the findings is based on differences between two datasets. Differences in SAR response quality between these sets are assumed to be caused by the introduction of the GDPR. These differences can also be caused by other differences in the two datasets, however. To minimize the risk of confusing differences in the dataset with influences of the GDPR, the data collection regime of the GDPR dataset was set up to mirror the data collection of the first set whenever possible. Aside from differences in the active regulation, the following differences remain:

- An intermediary method was used to send requests on behalf of volunteers during the GDPR data collection, while the wbp collection phase was performed on a personal basis.
- In using the intermediary method, emails were sent from the datarights.me platform, while the wbp collection phase was performed using personal and work-related email addresses
- Guidelines on sending reminders were stricter under the wbp data collection regime. Organizations responding to access requests in the collection of this set were therefore more often reminded of their duty to respond to the SAR than those contacted for the GDPR data collection.

To negate the effects of the first and second point, chapter 3 details a communication method that aims to catch and normalize responses of data controllers that "expose" the underlying research purposes of requests send for the GDPR data collection. Some of the data controllers that were asked to follow normal procedure after announcing the revelation of the SAR's research goals responded that it was hard to process a request made on behalf of a data subject in their normal process, since it was different from all requests that were seen before. All responding data controllers did say that answers on the requests would not be different than those that would be received with other request practices. Differences are therefore only expected to lie in

the additional time and number of responses required to check the validity in the method and to request for identification of two parties rather than one.

Others specifically stated the unusual datarights.me email address a feature that immediately raised red flags, since it was perceived to be an activist's website. This exceptionalism was also observed in other responses in the data collection process, where responses were more detailed than expected or send more often by DPO's or other specialized staff. Still, this difference in expected and received results was solely based on earlier responses under older regulation and thus no proof of verification problems. Furthermore, the data collection of the wbp dataset was performed mostly by researchers using their Delft University mailing address, which would have tainted result in a similar matter as using datarights.me addresses.

To verify how the effects of these first two observed differences in the dataset differ from the effect of the GDPR, a verification set is constructed that consist of data requests made personally by the researcher to organizations that are also included in the research. Although this set is too small for statistical comparison, the assumptions made about the previously identified differences seem to hold. Less correspondence was necessary to collect the verification dataset because the request method is never questioned, which also leads to quicker responses. The response rate and final response quality are similar if not the same, often having exactly the same format as was received during the intermediary data collection. One notable observation is a learning effect in two of the organizations which previously asked for excessive forms of identification. An alternative identification method that was agreed upon by both sides after a long discussion during the earlier intermediary data collection was immediately offered as an alternative in the verification request. This shows the impact that SARs may have on the privacy rights of others and some of the societal benefits that accompany this study's scientific contributions. This was also the conclusion of comparisons based on the second verification dataset, which includes responses made to organizations that were also contacted on behalf of other volunteers. When requests are sent on the same moment, responses will be the same (often exactly the same). When one is sent after the other, learning effects might have improved the process.

After a face validity check and testing the assumptions that follow from it using a control dataset, it is clear that analysis based on the number of correspondences and the total length of correspondence between SAR replies by organizations under the GDPR and wbp was too dependent on the different methods used. These two response factors are therefore not used in the quantitative analysis of SAR responses, but only in qualitative assessment of the data collection method

The third difference between datasets presents a bigger issue. Although both datasets stop their measurement after a total of 90 days, there was a difference in reminders sent to organizations. During the wbp data collection, a reminder was sent when the lawful deadline of four weeks for a response was surpassed. Because the novel collection method used to gather responses in the GDPR dataset and the single researcher that was available to send reminders, this lawful deadline was not strictly maintained. In the wbp data set, a total of 39 organizations received a reminder to respond to the request, with 21 organizations receiving a total of two reminders. In the GDPR set, only 9 organizations received a reminder. No second reminder were sent in the GDPR set.

In the earlier paper using the wbp dataset by Asghari et al. (2017) it is concluded that (more specific) reminders do lead to a higher response rate. It can thus be assumed that the response rate would have been higher under the GDPR if the same number of reminders was sent during the data collection phase. A higher response rate will also positively affect the other response variables measured in this research, since an organization can only be compliant or specific in its response once it has sent a response. It is therefore most likely that the effect of the GDPR on all other response scores is higher than was concluded in the quantitative analysis. However, the size of this effect is expected to be quite small, since the response rate was already quite high in the GDPR dataset. Furthermore, increases in the positive effect of the GDPR on SAR response quality does not change, but rather strengthens the conclusions that this positive effect does indeed exist. The difference in the number of reminders therefore only produces a minor limitation to the results.

With these tests, the internal validity of the model is established. This provides meaningful results and conclusions based on comparison between the two datasets. However, findings based on either of these datasets, or the generalization of results towards the real world should also be externally validated. This second validation investigates differences between findings as observed in the analyzed sample and the existence of these findings in the general population.

The first potential limitation in the generalization of results is based on the organizations that are included in the sample that is analyzed. For the GDPR dataset, this sample is very dependent on the organizations that volunteers are willing to include in the study. This choice is further limited by the organizations that can be reasonably expected to process the volunteers' personal data. In annex A.5.3, the characteristics of volunteers, and the organizations that these volunteers shared with the researcher are analyzed. As was already anticipated in the demarcation of the scope in chapter 3, most volunteers participating in the research are students and under 30 years old. This has not limited the choice of organizations for the research however, since most volunteers were eager to suggest many organizations for research during the kick-off meetings. The generalization of results is thus not limited by the number of organizations available to sample.

A second challenge in generalizing results based on research like this is that the data collection is often performed by data privacy experts, while the conclusions are generalized to all data subjects (Norris et al., 2017). Although the researcher conducting was certainly no expert in the GDPR and the (other) specific regulation and exemptions concerning SARs at the start of the research, the large amount of developed experience certainly improved the ability to more efficiently counter certain unsatisfactory reply. Most of these encounters however, consisted of responses questioning the intermediary method used. Furthermore, no resources were used that were no available to normal data subjects, with most originating from the DDPO and ICO website. It can therefore be confidently stated that the effect of this bias did not significantly affect findings. If a positive effect of "expert" responses does exist, it is very likely to be smaller than the negative effect on results that follows from the lenient reminder regime. Results are thus believed to be externally valid.

5 Discussion

The previous chapter has provided insights into organization's reactions to Subject Access Requests (SARs) in the dataset in general, with a specific emphasis on the factors that seem to influence these reactions. This discussion chapter will put the findings in a societal perspective and detail the challenges that remain. The chapter is not only based on the researcher's findings and experiences, but also incorporates the opinions of some of the volunteers and data controllers involved during the data collection and research phases of this thesis research.

5.1 Access rights under the GDPR

Earlier findings provide a perspective of exercising privacy rights which is challenging at times but has certainly improved a lot compared to earlier implementations in European law. This is perhaps best explained by comparing the results of the previous section directly to those found in previous research. Table 14 compares the results to earlier research referenced in chapter 2. The comparison clearly shows that the improved positive outcome is not the result of a larger share of responses, but rather by the quality of these responses. This is especially evident in the responses to further questions regarding the processing of personal data. The GDPR has thus caused a real culture change in the way organizations handle access requests. This culture change is also visible in the processes that are in place in organizations processing requests, as detailed in the qualitative results of the previous chapter.

Study	N	Country	Sectors	Response rate (1)	Response with data (2)	Response with answers (2)	Compliance
Norris et al., 2017	183	EU	Mixed	80%	57%	43%	34%
Asghari et al., 2017	106	The Netherlands	Mixed	83%	69%	27%	22%
Herrmann & Lindemann, 2016	120	Germany	Popular apps & websites	68%	n.a.	n.a.	43%
Spiller, 2016	17	UK	CCTV	n.a.	35%	n.a.	35%
Ausloos & Dewitte, 2018	60	EU	ISSS	74%	67%	n.a.	(3) 33%
GDPR part van Biemen, 2018	116	The Netherlands	Mixed	81%	68%	(4)65%	51%

Table 14. *Previous results of research into organizational responses to subject access requests.*
(1) The response rate includes responses received after the legal deadline and/or received after multiple reminders, which can be considered non-compliant to the law. (2) Data that was deemed incomplete or inaccurate are excluded, when detailed by the researchers' results. Responses without data can still be considered as a good response in this category if participants believed that the organization was indeed not or no longer in possession of their personal data. (3) Based on satisfaction rating by participants, as legal outcome was not reported. (4) Organizations that claimed to not have any data on the subject and thus were not obliged to answer further questions are also included in this measure.

During the data collection phase of this research, some organizations indicated that they were still working on finalizing their process to be fully compliant or more facilitative in responding to SARs. The effect that this will have on the overall compliance rate is expected to be quite small however, since these organizations often invested extra time to respond in compliant ways without having all processes up and running. Only a few organizations responded in a non-compliant way that was clearly because of easily tweaked internal processes. The one-time boost that the introduction of the GDPR has provided is therefore not expected to be increased in the following months without further actions from regulators and policy makers. With the increased power that the GDPR has vested into regulators, and the clean starting point it has provided (national) lawmakers, a wide range of tools is available for them to increase the effectiveness of personal privacy regulation further. Following findings from this research, proper governance should attempt to balance the anxiety within and between organizations with regards to privacy regulation. The objective of this balance is best explained through the introduction of two elements of anxiety that were observed in or communicated by organizations approached during this research.

The first dimension of organization's anxiety follows from consequences in infringing the GDPR. The introduction of higher fines, wider jurisdictions and more efficient European collaboration of regulators increases organizations' perceived risk of penalties in not complying with privacy regulation. This develops the once harmless regulators into a force perceived to be very dangerous to businesses, which in turn ensures that data protection is a higher priority for any organizations. For strong data subject's privacy rights, it is vital that data protection will remain a priority for organizations. This is only possible when lowering this priority continues to be (perceived as) a risk. A certain level of anxiety over these risks within organizations is therefore vital to ensure peoples customer privacy rights. In the case of SARs, it is also important that data subject keep exercising their personal rights in order to resume organizations anxiety over their importance and reality under the GDPR. Furthermore, it will ensure regulator receive a steady stream of insights into which organizations seem to be non-compliant.

Findings in this research show that high GDPR anxiety may also decrease the insight that data subjects can obtain using SARs as organizations are too anxious to release the required information. This second observed dimension of anxiety often follows from a misunderstanding of the GDPR. Most of the non-compliant organizations that have been identified in this research were unknowingly in breach of the GDPR. These organizations often followed the strictest protocols applicable in order to rule out noncompliance of other GDPR passages. In interviews with data controllers, it is clear that the transparency of information that should be provided through SAR answers was less important to them than the secure protection of this information within the organization. Some organizations prefer to keep information regarding certain data processing practices to themselves in order to ensure security through security.

An example of this second GDPR anxiety dimension is found in the identification phase of SARs. Some organizations declare that the conventional SAR method of identification, providing a copy of an identity card, is not enough to verify a person's identity in the degree that is required under the GDPR. Although a copy of an identity card certainly does not provide solid proof of identity, these organizations often interpret their identity verification duty far too strict. In looking for other means of verification, organization often dismiss the very method that was used in obtaining the personal in the first place. These data controllers also dismiss the option to use the data in their possession to verify the data subject's identity, since this is considered as unlawful processing. Through misinterpretation of regulation, these organizations often require excessive amounts of new personal information in order to identify the person exercising their privacy right, thereby breaking the very GDPR they are trying to follow.

Although the DPA provides data subjects with clear examples on how to enforce their rights, data controllers are supplied with long, technical documents on specific issues. If organizations do not have a good example of a verification method that is both easy to use and GDPR compliant, they are left to their own devices to design

such a process. In combination of the first anxiety dimension, this will lead to a rigid and harsh process which will discourage the facilitation of transparency.

Too little of the first dimension of GDPR anxiety will lead to even more widespread non-compliance, which makes the regulation unenforceable in practice and remove personal data rights from mainstream practice again. Too much of the second anxiety dimensions will have the same effect, because organizations will fail to comply with regulation that is unclear or in their eyes even internally contradictory. It is no surprise that data controllers react defensively in these cases by releasing as little information as possible until it is certain that releasing more information is indeed required. In order to save the reclaimed SAR rights from becoming unenforceable, it is thus important that the correct balance of anxiety from enforcement and anxiety from misunderstanding is found. The introduction of the GDPR, and the hype that surrounded it, has provided regulators with a window of opportunity to improve this balance. The following policy recommendations could help achieve the desirable balance:

- The DPA should fine organizations that are not compliant with the GDPR before their anxiety window of opportunity closes. In fining organizations, special attention should be payed towards to the data controller's intention in infringing on the GDPR.
- The DPA and government should ensure that data subjects will keep sending SARs. This can either be achieved through popularity campaigns or by creating a platform to exercise their rights collectively.
- The DPA or government should clearly state that current practice of restrictive SAR tools and excessive identification requirements are in fact in breach of the GDPR, as these limit data subjects in achieving their rightful insight.
- The DPA or government should provide data controllers with GDPR compliant tools or advice on ways to better check the subject's identity, since a copy of one's identity card can easily be tempered with. One improvement opportunity may be found in the expansion of the Dutch governments DigID project, or the acceptance of similar digital identifiers.
- The European DPA should also offer a clarification statement explaining their advice on what is meant with the use of "all reasonable measures to verify the identity of a data subject who requests access". (GDPR rec. 64)
- The European DPA should provide best practices and examples of responses to access requests that are both fully compliant and very facilitative in their presentation. Ideally, these best practices should also show the process that was carried out internally within a compliant organization.

5.2 Access through Procurement

This research has presented a novel method of data collection by exercising access requests on behalf of volunteers. Data creation in this way was inspired by mass collaboration methods as described by Salganik (2018). The method, which is described in detail in chapter 3, aimed to allow the faster creation of a larger and more focused dataset. The method, as it was implemented in this research, did succeed in creating a larger dataset, which was more focused toward organizations that were deemed interesting for research than was possible with only requesting access to the researcher's organizations. It did however also present certain challenges that made the data collection slower rather than faster.

One of these challenges was found in data controllers' unawareness of its legality. Many data controllers were not easily convinced of the lawfulness of the data collection method, even when specific quotes from the Dutch law code were presented as proof for its legality. This extra process made exercising data rights significantly more cumbersome. Convincing organizations of the lawfulness of access rights through procurement is expected to be easier in the future, as the Dutch and UK DPA's now both have clearly stated on their website that this method is indeed a valid one for exercising data rights.

A second challenge was found in the division of labor. Mass collaboration works best if work is divided equally among participants. This was not the case during the data collection of this research, in which almost all responses to data collection followed from conversations made by the researcher. Further implementations of this method should strive to recruit volunteers that are highly motivated in researching the processing of their own data too. Another possible point of improvement would be the introduction a gaming element using of some kind of scoring system for actively exercising requests, with researchers awarding points to volunteers that obtain responses to access requests on the basis of their data's usefulness for the research. This element is also something that is found to be important for other mass collaboration examples by Salganik (2017). However, researchers should not be too detached from the collection in practice, as the process itself can provide them with important details to explain certain trends in the dataset.

Although the implementation was certainly challenging at times, the use of this novel method has certainly helped to deliver insight into SAR responses and the factors that affect these responses in a way that could not have been achieved on this scale without the help of volunteers. The insights of this research thereby show the upsides of a method with huge potential, when implemented correctly. The method also has the potential to contribute to data subject's personal privacy rights beyond scientific contributions. A direct societal contribution found in the data collection method was that it enabled those with little ICT and/or data privacy experience to still exercise their privacy rights effectively. Researchers in this way can act as consultants in advising volunteers on the best way to respond to certain responses. When necessary, a researcher can also use his expertise to deal with specifically tricky responses on behalf of the volunteer.

5.3 Limitations and further research

In interpreting the results and conclusions following this research, it is important to note the existence of some limitations. One limitation was found in results following the comparison of SAR responses of organizations responding under the wbp and GDPR jurisdiction. Although researchers have tried their best to reduce the differences between data collection under the two jurisdictions as much as possible, a difference was still found after testing for validation. This difference lies in the number of reminders that were sent to organizations that did not immediately respond to SARs in the collection of the two underlying datasets. The difference negatively affected the response rate measured under the GDPR. Since this response rate influences organization's response scores both directly and indirectly, compliance and specificity rates of GDPR bound organizations are probably reported lower than they actually are. The difference is not assumed to be very high though, since the response rate is already quite high and on a comparative level with earlier research.

The second limitation on conclusions presented based on this research is that data collection is performed mostly by researchers with a growing degree of experience in personal privacy rights, and access rights in particular. This increases the risk that certain organizations would have responded differently to less experienced data subjects. However, as this expertise was only perceived to be useful during data collection in response to data controllers rejecting SAR's based on the procurement method that was used, it is not expected to have a large effect. This situation is not expected to be a problem to the average data subject, even less so now the Dutch DPA has explicitly stated the legality of the method.

One final limitation of this research lies in the validation method itself. In the absence of a larger validation dataset, this validation is completely based on qualitative analysis. And although this analysis should (and did) catch big validity problems, it does not carry the thoroughness of qualitative validation methods. Following this limitation, it is advised that further research focuses on numerically validating the data collection method that was used for this research. Through this validation, the insights presented in this thesis report would become even more valuable. Furthermore, it would also validate the dataset that has led to these insights.

Because of the cumbersome data collection, there was not enough time to analyses all potential relationships that could be found within the data. Further research could, for example, use the (anonymized) dataset to investigate the impact of organizational variables on specific result variables such as the response rate. Another possible research topic would be to investigate the increased compliance that the GDPR has caused in different sectors. Analysis of this difference in increased compliance could also be interesting for organizations of different sizes, since the GDPR's mandatory DPO is expected to have at least some influence on big organizations' compliance levels. Further research opportunities originating from un-analyzed variables in the dataset is a possible relationship between volunteers' characteristics or perceived skills and organizations SAR responses. The dataset also contains variables that are not used numerically in this research in the mean response time, number of discourses and communication methods of organizations.

The findings of this thesis research also provide a fair amount of further research opportunities. Since these results are largely based on data that was collected just after the GDPR became enforceable, it might be interesting to see how compliance, presence and nature of certain practices evolves through time. The results from this research can also be used as a baseline for research into the influence of certain other factors, such as newly introduced policy. Further research into responses made by organizations in other European countries could measure the impact of culture on these responses. Another opportunity lies in research concerning the relationship between factors influencing organization's compliance of to the right of access and their compliance in other personal privacy rights, or the GDPR at large.

The final research opportunity that the researcher would like to be pursued is the improvement of the procurement method for data collection. Although the method has certainly made data collection for this thesis research more cumbersome and time consuming, it has also shown real potential for creating bigger and more focused dataset more quickly. Improvements in the division of labor as suggested in section 5.2 can provide a basis for improvements.

6 Conclusions

Citizen's fundamental rights to informational self-determination has been a cornerstone of European data privacy regulation for almost forty decades. However, research into the exercising of this right in practice has been sparse for many of those years. The use of Subject's Access Request (SAR), which acts as a natural precursor to exercising the other rights that should guarantee informational self-determination and data privacy, has existed in an investigatory vacuum until publications by Norris et al. (2017). Research has since shown a great disparity between the European data privacy law in writing and the application of the law in practice. However, research has not conclusively shown what factors explain the way in which organizations respond to SARs.

The introduction of the GDPR provides a new opportunity to understand these factors and their relationship with responses to access requests. This research aims to provide this insight by combining data that was used to investigate SAR responses under old jurisdiction with new data collected using a novel intermediary approach to sample SAR responses under the GDPR. Previous sections present the results from both qualitative and quantitative analysis of differences in these responses. These findings, supplemented with insights into further societal consequences in the previous section, allow for the first conclusive insights into the effect that the GDPR and other factors have on organizations in their answer to access requests.

6.1 Main research question

How are organizations responding to Subject Access requests under the GDPR, and what factors influence their responses?

Out of a sample of 116 Dutch organizations, 51% responded to subject access requests (SARs) in a manner compliant to the GDPR. Although this means that non-compliance is still widespread, the share of compliant organizations is significantly higher than in any similar research performed under previous regulation. Regression analysis shows that the introduction of the GDPR has at least doubled and most likely tripled the chance of a compliant answer. This effect might even be higher, because of differences in the collection regimes of the underlying datasets. Qualitative analysis of SAR responses from a further 249 organizations affirms this conclusion. Although a large group of organizations still does not provide the required insight into the processing of personal data, the GDPR has clearly increased the responses of subjects' access requests. The regulation has thus provided European citizens with a more effective tool to guarantee their right to informational self-determination.

For many organizations, access requests seem to no longer be an obscure incident, but a real possibility. This can probably be attributed to the data protection hype that surrounded the introduction of the GDPR. And although this has also seemed to improve the general attitude of organizations responding to such requests, it also causes adverse effects on organization's responses. These adverse effects are most apparent in the demands that are made by some data processers for subjects to proof their identity beyond reasonable means. Another adverse effect is found in restrictive processes that many organizations have created to deal with SARs in a more efficient way. By streamlining requests, these processes often only provide partial insight into the processing of personal information that data subjects are lawfully required to receive.

Apart from the introduced GDPR, the only factor shown to have a clear quantitative influence on SAR responses is the size of an organization. Smaller organizations produce better responses to access requests, with a difference that is comparable to the effect of the GDPR introduction. This is because smaller organizations are more prone to answer SAR's in a more personal and specific matter. No sectors are found to be significantly less or more compliant in their responses, although the ICT and other services sector are shown to be more prone to answer requests in a non-specific manner. Conflicting regulation in the healthcare

sector and communication practices in the public sector does create specific hurdles to receive insight in personal data processing of these sectors. Other investigated effects are not observed to influence organization's reported SAR response results.

6.2 Research subquestions

Following the introduction of a conceptual model of factors that may influence the response of data controllers to SARs, four research subquestions are drafted in section 2 to investigate the effect of four factors groups expected to be the most influential. Hypothesis on the effect of these judicial, organizational, subject and request factors were tested using mixed method research in subsequent sections. Findings concerning the influence of each factor group is combined in this section to answer each of the four subquestions specifically.

1. *What judicial factors influence data controllers' responses to subject access requests, and to what extend?*

Although many organizations are still not able to handle SARs correctly, the introduction of the GDPR is shown to have a largely positive impact on organizations' responses to such requests. Regression analysis in section 4.2 shows organizations are at least 64% and at most 600% more likely to respond in a compliant manner under the new regulation. This effect might be even higher in reality because of a difference in data collection regimes. Further qualitative analysis also shows an improvement in the way organizations respond after introduction of the GDPR. Data subjects' access rights are more often acknowledged, and responses are often more complete. The introduction of a mandatory DPO in larger organizations also seems to have improved the SAR process by increasing data subjects' power to raise issues with incorrect responses.

The introduction of the GDPR has removed a big barrier that was holding back the use of SARs in practice: it's obscurity. It has however also led to the creation of two new barriers. The first is found in processes which are designed by predominantly larger organizations to deal with requests in a more streamlined way. These processes can confine data subjects in their requests and often only lead to partial insight into the processing of their personal data. The second barrier is found in excessive identification methods that are imposed by data controllers. Controllers often place large burdens on subjects to proof their identity, thereby requiring the processing of even more personal data. These barriers are seen as a symptom from a selective hype that surrounded the GDPR implementation. Organizations seem anxious to protect the personal data in their possession but fail to see how their practices infringe on other privacy and GDPR fundamentals such as transparency and data minimalization.

2. *What organizational factors influence data controllers' responses to subject's access request, and to what extend?*

The way organizations react to access requests is also influenced by characteristics of these organizations. The most influential of these characteristics is their size. Organizations with fewer than 25 employees are far more likely to be compliant in their SAR response than bigger counterparts. This effect contradicts the assumption that SAR response compliance is a matter of resources and challenges the idea that the stricter GDPR is a bigger liability for smaller organizations. Further analysis shows that the effect of organizations' size on SAR responses can be explained by the more specific, personal approach that is taken by smaller organizations in dealing with SARs.

Although differences in reaction also seem apparent in different sectors, no clear trend is found. Numerical analysis has also not shown a significant difference in the compliance rates of organizations in different sectors. Differences in the specificity of organizations' responses is found, with organizations in the ICT and other services sectors responding in a noticeably more generic way to SARs compared to other organizations.

Qualitative analysis shows differences in responses from organizations in two other sectors. Governmental organizations show a distaste for lawfully required electronic communication methods which would ease the process of requesting personal data. Organizations in the Dutch healthcare sector follow stricter and sometimes conflicting sector specific regulation which limits data subject's right to access certain information.

Clear differences in compliance or response practices between national, European Union and international organizations were not found in both qualitative and quantitative analysis. This is probably because of the small number of non-Dutch organizations in the sample but can also be explained by the correlation between an organizations size and operating location.

3. *What subject factors influence data controllers' responses to subject's access request, and to what extend?*

Volunteers were very reluctant to include organizations with which they had a more personal relationship in this research. Almost all of those that were added, were withdrawn in later stages. Although these choices show the importance of relationships with organizations concerning SARs, research into their influence in data controller's answers remains inconclusive. Further qualitative analysis has also not yielded

4. *What request factors influence data controllers' responses to subject's access request, and to what extend?*

Although this could not be verified in quantitative fashion, differences in the request itself do seem to influence an organization's response. The intermediary method that was used in this research: sending SARs on behalf of others, made organizations more anxious to share personal data, even when the process was shown to be grounded in Dutch law. Qualitative investigation of organizations' responses shows that this made the SAR process longer and harder, because of an extra barrier that researchers had to overcome in proving the legality of the request. The legal time requirement for SAR responses was excluded in the judgement of compliance to compensate for the extension that this brought to the data collection process by organizations that often responded quickly after the extra barrier was overcome. Because of this measure, the difference in request form is not expected to have influenced the other conclusions on organizations' reactions to SARs.

6.3 Policy recommendations

Data subjects' informational self-determination through SARs has clearly improved with the introduction of the GDPR. The one-time boost in data privacy awareness that it caused in organizations also provides policymakers with a window of opportunity to establish a long-term obligation to personal privacy rights. In chapter 5, policymaker's challenge to utilize this opportunity is analyzed as a balance of anxiety dimensions. A certain degree of anxiety following from the first dimension, the perceived risks of infringing on the new regulation is necessary to ensure the priority of properly handling SARs by organizations. This should be balanced with a second dimension of anxiety following from misunderstanding and ambiguity within the regulation, since this uncertainty discourages organizations to facilitate transparency.

In this research, two big hurdles in exercises SAR's have been identified. Both follow from an improper balance of anxiety and limit both data subjects and controllers in the access request process. The first of these is found in excessive identity checks that data controllers impose on subjects exercising their data rights. The following policies are designed to dismantle this barrier:

- The DPA or government should clearly communicate to organizations that data subjects should not have to provide personal information that is not currently in the controllers' possession to proof their

identity. This includes organizations that compel subjects to proof their identity in person when less privacy infringing methods are available.
- The DPA or government should provide data controllers with examples of GDPR compliant tools or advice on ways to check the subject's identity without further privacy infringement. The copy of one's identity card, which is currently advised as proof, does not provide the controller with a high level of verification. Improvement opportunities can be found in the expansion of the Dutch government's DigID project, or the introduction of similar, trusted digital identifiers.
- The European DPA should offer a clarification statement explaining their advice on and limits to what is meant with the use of "all reasonable measures to verify the identity of a data subject who requests access" (GDPR rec. 64).

The second hurdle is found in restrictive processes implemented by data processors. These processes limit data subjects to just partial insights into the processing of their data. The following policies are designed to dismantle this barrier:

- The DPA should make sure organizations know the full extent of data subject rights in the GDPR, which is far broader than a plain copy of processed personal data.
- The DPA should advice organizations on how to respond to SARs on a more specific level. Ideally, best practices are shared on the DPA website, so that organizations have an example to compare their process to. Nowadays, most information that the DPA has made available is geared towards the data subjects, which leaves data controllers without reliable examples.
- The government or DPA introduces a simple benchmark to check if SAR processes facilitate the request of all information that data subjects should be able to request.

The required anxiety balance also requires continual effort of the DPA on longer terms. In balancing the first dimension, it is important that the DPA shows that violations of the GDPR will indeed be punished. In their penalty assessment, the DPA should also consider what motivated the data controller in their error. The divisions in terms of compliance and specificity in chapter 4 can be a starting point for dividing organizations not only based on outcomes, but also differentiate between those that are somewhat compliant an those that are not at all. The recent penalty that the Dutch DPA imposed on a bank which did not comply with an earlier access request and the inclusion of a wide arrange of motivational factors available on deciding the size of such penalties in article 83 of the GDPR suggests a good starting point for the balance in anxiety based on GDPR infringement. In balancing the second anxiety dimension, it is vital that uncertainty over parts of the GDPR are constantly addressed. Especially when ambiguity is not yet addressed in court, the DPA should publish their view on the matter. After punishing violators, the DPA should detail what the penalized party should have done better, so that others can learn from their mistakes.

A final steering role for governments can be found in improving sector specific regulation that interferes with the GDPR, as was found in the Dutch healthcare sector. In GDPR compliance, the government should set a better example. Organizations in the public sector for example, own all tools required to design an optimal SAR experience, including full disclosure of personal information, electronic communication of results and the use of digital identifiers. Instead, most local governments have opted for paper, restricting processes and obligatory identity checks in person.

6.4 Scientific contribution

Aside from the insights into the measures that should be taken to further help privacy rights, this research provides multiple scientific contributions. The first of these contributions is in filling the knowledge gap that was identified in section 2, since this research provides the first conclusive insights into the impact of the GDPR and the other factors that influence exercising access rights in practice. In doing so, a second scientific

contribution is found in the conceptual model of section 2.3.1. This is the first attempt at modelling the factors that influence SAR's, incorporating both the individual and societal perspective of their application.

Two further scientific contributions follow from the data collected for this research. The first of these is the dataset that formed the basis of the regression analysis of section 3.1. This data is expected to be able to deliver further insight into the practicalities surrounding exercised SARs, as detailed in the previous chapter. The final and arguably important scientific contribution of this research lies in the novel data collection method that was used. Performing access requests on behalf of others is a method that has not been used before in research but is shown to be both feasible and helpful in sampling data that is out of reach of the researcher in personal SARs. Furthermore, the detailed description of the application of the method for this research in chapter 3, combined with the proposed improvements in chapter 5, provide researchers with the tools to further develop this method to one that takes advantage of all benefits mass collaboration has to offer.

6.5 Societal contributions

As an EPA thesis, this research's scientific insights are certainly supplemented with societal contributions. The first contribution is found in the central factor of this research, the GDPR. Many opinion pieces have been written on the effect of certain (omitted) passages after their introduction. This research however, provides the first scientific contribution on its effects in practice, thereby providing the first set of measurable facts to advance the discussion on the effects of what some deem to be the most contested law in the EU's history (Powles, 2018). More specifically, this research provides policy makers with a first analysis on the hurdles that still exist in exercising SARs. These contributions are supplemented by policy recommendations stemming from a combined researcher, data subject and data controller perspective, an important third societal contribution.

The final societal contribution that follows from this research is the practice of performing access requests on behalf of others. As discussed in the previous chapter, this method has the potential to provides even those without any understanding of both data privacy and the surrounding regulation with their rightful insights. This research has set an important and sometimes challenging first step in using this method in practice. The decision of the Dutch DPO to include this method as a possibility on their guide on exercising personal privacy is expected to increase both the awareness and convenience of helping others in pursuing their informational self-determination. In the author's view, this decision serves as an acknowledgement of this most important societal contribution.

Bibliography

Alaqra, A. S. (2018). *The Wicked Problem of Privacy: Design Challenge for Crypto-based Solutions.* (Dictoral dissertation, Karstads universitet).

Albrecht, J. P. (2016). How the GDPR will change the world. *Eurpean Data Protection Law Review*, 2(3), 287.

Allen, D., Berg, A., Berg, C., Markey-Towler, B., & Potts, J. (2018). Some Economic Consequences of the GDPR. *SSRN Electronic Journal.*

Asghari, H. (2016). *Cybersecurity via Intermediaries: Analyzing Security Measurements to Understand Intermediary Incentives and Inform Public Policy* (Doctoral dissertation). Retrieved from https://repository.tudelft.nl/islandora/object/uuid:3694edf5-d6e0-4484-b847-750da2b9d1b9

Asghari, H., Mahieu, R. L. P., & van Eeten, M. (2017). *Collectively Exercizing the Right of Access: Individual Effort Societal Effect*. Paper presented at GigaNet Annual Symposium, Genova. Retrieved from https://hcommons.org/deposits/item/hc:17975/

Assembly, U. N. G. (1948). Universal declaration of human rights. "Universal Declaration of Human Rights." *United Nations, 217*. Retrieved from http://www.verklaringwarenatuur.org/Downloads_files/Universal%20Declaration%20of%20Human%20Rights.pdf

Ausloos, J., & Dewitte, P. (2018). Shattering one-way mirrors – data subject access rights in practice. *International Data Privacy Law, 8*(1), 4–28.

Autoriteit Persoonsgegevens (2018, June 29). *Ruim 600 mensen dienen privacyklacht in bij AP*. Retrieved from https://autoriteitpersoonsgegevens.nl/nl/nieuws/ruim-600-mensen-dienen-privacyklacht-bij-ap?_sp=84e6896f-1dbe-4a3c-a0be-a5fd1c443ce4.1535232139428

Autoriteit Persoonsgegevens (2018b) *Recht op inzage.* Retrieved from https://autoriteitpersoonsgegevens.nl/nl/zelf-doen/privacyrechten/recht-op-inzage

Axinte, S.-D., Petricã, G., & Bacivarov, I. (2018). GDPR Impact on organziation Management and Processed Data. *Quality-Access to Success, 19*(165). 150-153

Balci, O. (1994) *Validation, verification, and testing techniques throughout the life cycle of a simulation study*, Annals of Operations Research, 53: 121. https://doi.org/10.1007/BF02136828

BBC News. (2017). *Is privacy dead in an online world?* Retrieved from https://www.bbc.com/news/technology-41483723

Beach, F. A., & Diamond, M. (1977*). Human sexuality in four perspectives.* Baltimore: Johns Hopkins University Press.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., Venkatraman, N. (2013). Digital Business Strategy: Toward a Next Generation of Insights. *The Mississippi Quarterly, 37*(2), 471–482.

Bijron, Z. (2017, July 6). *Hoe geef je invulling aan de verschillende rechten van betrokkenen: recht van inzage en recht op rectificatie*. Retrieved from https://www.privacycompany.eu/het-recht-van-inzage-en-het-recht-op-rectificatie-een-praktische-invulling/

Birds, T. (2017). *Guide to the General Data Protection Regulation*. Retrieved from https://www.twobirds.com/~/media/pdfs/gdpr-pdfs/bird--bird--guide-to-the-general-data-protection-regulation.pdf

Birnhack, M. D. (2008). The EU Data Protection Directive: An engine of a global regime. *Computer Law & Security Review, 24*(6), 508–520.

Birnhack, M., & Elkin-Koren, N. (2010). Does Law Matter Online-Empirical Evidence on Privacy Law Compliance. *Michigan Telecommunications and Technology Law Review, 17,* 337-384.

Blakey, G. R. (1964). The Rule of Announcement and Unlawful Entry: Miller v. United States and Ker v. California. *University of Pennsylvania Law Review, 112*(4), 499–562.

BOF. (n.d.). *PIM stopt er mee*. Retrieved from https://pim.bof.nl/

Borrás, S., & Edler, J. (2015). *Governance of Socio-Technical Systems : Explaining Change.* Cheltenham: Edward Elgar Publishing, Incorporated.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society, 15*(5), 662–679.

Bradford, A. (2012). The brussels effect. *Northwestern University Law Review, 107*(1), 1-68.
Branscomb, A. W., & Larson, J. (1995). Who owns information? From privacy to public access*. College and Research Libraries, 56*(2), 186–187.
Brill, J. (2018, May 21). *Microsoft's commitment to GDPR, privacy and putting customers in control of their own data*. Retrieved from https://blogs.microsoft.com/on-the-issues/2018/05/21/microsofts-commitment-to-gdpr-privacy-and-putting-customers-in-control-of-their-own-data/
Burgerlijk Wetboek Boek 3 - BWBR0005291. (1992, January 1). Retrieved from http://wetten.overheid.nl/BWBR0005291/2017-09-01
Bu-Pasha, S. (2017). Cross-border issues under EU data protection law with regards to personal data protection. *Information & Communications Technology Law, 26*(3), 213–228.
Busch, D. Hondius, E., Van Kooten, H., Schelhaas, H., & Schrama, W. (2003). The Principles of European Contract Law and Dutch Law, A Commentary. *Revue Internationale de Droit Comparé, 55*(3), 706–708.
Cagnin, C., Amanatidou, E., & Keenan, M. (2012). Orienting European innovation systems towards grand challenges and the roles that FTA can play. *Science & Public Policy, 39*(2), 140–152.
Cauer, E., Mathis, W., & Pauli, R. ,(2008) Life and Work of Wilhelm Cauer (1900 – 1945). *MTNS2000, 1-10.*
CBS (2008) *Standard Industrial Classifications (Dutch SBI 2008, NACE and ISIC)* Retrieved from https://www.cbs.nl/en-gb/our-services/methods/classifications/activiteiten/standard-industrial-classifications--dutch-sbi-2008-nace-and-isic--
Cellan-Jones, R. (2018, May 25). The great GDPR panic*. BBC.* Retrieved from https://www.bbc.com/news/technology-44240664
Charlesworth, A. (n.d.). *A Very Short History of Data Protection*. Retrieved from http://www.cloudview.co/Averyshorthistoryofdataprotection
Chiavetta, R. (2017) *Survey: 61 percent of companies have not started GDPR implementation*. Retrieved from https://iapp.org/news/a/survey-61-percent-of-companies-have-not-started-gdpr-implementation/
Clarip (n.d.). *Data Subject Access Request (or Rights) DSAR Portal for GDPR Compliance Software*. Retrieved from https://www.clarip.com/data-privacy/dsar-portal/
Council of Europe. (1981). Convention for the Protection of Individuals with regard to Automatic Harford, T. (2014). Big data: A big mistake Processing of Personal Data Retrieved August 19, 2018, from https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680078b37
Council of Europe. (1950). *Details of Treaty No.005*. Retrieved from https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005
Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks: Sage Publications.
Curtis, J. (2018, April 27). *Meeting the GDPR deadline: Don't panic, and show your working.* Retrieved from http://www.itpro.co.uk/general-data-protection-regulation-gdpr/31009/meeting-the-gdpr-deadline-dont-panic-and-show-your
Custers, B., Dechesne, F., Sears, A. M., Tani, T., & van der Hof, S. (2018). A comparison of data protection legislation and policies across the EU. *Computer Law & Security Review, 34(*2), 234–243.
datarights.me. (n.d.). Retrieved from https://datarights.me/
Davies, J. (2018, May 25). *GDPR mayhem: Programmatic ad buying plummets in Europe - Digiday*. Retrieved from https://digiday.com/media/gdpr-mayhem-programmatic-ad-buying-plummets-europe/
Davis, J. S., & Osoba, O. (2018). Improving privacy preservation policy in the modern information age. *Health and Technology, 8*(4), 1-11.
Davison, R. M. (2007). *The New Economy in Development: ICT Challenges and Opportunities. The Electronic Journal of Information Systems in Developing Countries, 29(*1), 1–1.
DeCew, J. (2018). Privacy. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2018). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2018/entries/privacy/
Dehaye, P. O. (n.d.). *About*. Retrieved from https://personaldata.io/about/
Dehaye, P. O., Hahn, I., & Jargalsaikhan, G. (n.d.). *Platforms and personal data processing: the potential for achieving systemic transparency.* Retrieved from

https://www.dropbox.com/s/0c9xh10oek2jl1s/Platforms%20and%20data%20subjects%27%20rights.pdf?dl=0
Denhart, C. A. (2018, May 25). New European Union Data Law GDPR Impacts Are Felt By Largest Companies: Google, Facebook. *Forbes Magazine.* Retrieved from https://www.forbes.com/sites/chrisdenhart/2018/05/25/new-european-union-data-law-gdpr-impacts-are-felt-by-largest-companies-google-facebook/
Dine, J., Koutsias, M., & Blecher, M. (2007). *Company Law in the New Europe: The EU Acquis, Comparative Methodology and Model Law*. Cheltenham Glos: Edward Elgar Publishing.
European Data Protection Supervisor (n.d.). *A - European Data Protection Supervisor*. Retrieved from https://edps.europa.eu/data-protection/data-protection/glossary/a_en
European Data Protection Supervisor (n.d.). *The History of the General Data Protection Regulation - European Data Protection Supervisor.* Retrieved from https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en
European Parliament (1995). *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*. Retrieved from http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046
Esteve, A. (2017). The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law, 7*(1), 36–47.
Eurostat (2008). *Statistical Classification of Economic Activities in the European Community, Rev. 2.* Retrieved from http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIC&CFID=1505519&CFTOKEN=72ade128432278a0-8C677D5A-BCBE-A71A-453E62B48EF6DCAB&jsessionid=ee30c97ea7f3644f7a
European Commission (2016). *Europe 2020 strategy*. Retrieved from https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester/framework/europe-2020-strategy_en
European Commission (2018). *It's your data – take control – Data protection in the EU.* Retrieved from https://ec.europa.eu/commission/sites/beta-political/files/data-protection-overview-citizens_en.pdf
EY. (2015). *Innovating with RegTech, Turning regulatory compliance into a competitive advantage.* Retrieved from https://www.ey.com/Publication/vwLUAssets/EY-Innovating-with-RegTech/%24FILE/EY-Innovating-with-RegTech.pdf
Faradina, D. A. R., (2017). *Towards the Adoption of EU General Data Protection Regulation: An Empirical Study of Businesses' Perception on Privacy and Data Protection* (Master's thesis). Retrieved from http://resolver.tudelft.nl/uuid:bdc4de27-5c2e-4d13-8979-4a3c75e78bfd
Financial times (2017) Data protection: Brussels' heavy hand on Europe's digital economy. *Financial Times*. Retrieved from https://www.ft.com/content/777a1d34-ceb4-11e7-b781-794ce08b24dc?segmentId=9b41d47b-8acb-fadb-7c70-37ee589b60ab
Fiore (2018). *GPs asked to lobby MPs over "bombardment" of patient data requests.* Retrieved from http://www.pulsetoday.co.uk/news/gp-topics/it/gps-asked-to-lobby-mps-over-bombardment-of-patient-data-requests/20037259.article
Folsom R., Lake, R. B., Nanda, V. P. (1996). *European Union Law After Maastricht: Practical Guide for Lawyers Outside the Common Market*. Den Haag: Kluwer Law International.
Ford, C. S., & Beach, F. A. (1951). *Patterns of sexual behavior*. New York: Harper and Row.
Fox, T. (2018, May 22). Countdown to GDPR: Subject Access Requests (SARs). *JDSupra.* Retrieved from https://www.jdsupra.com/legalnews/countdown-to-gdpr-subject-access-85079/
Franceschi-Bicchierai, L. (2014). The 10 biggest revelations from Edward Snowden's leaks. *Journal of Law and Technology at Texas*, *1*, 1-217.
Frowein, J. A. (n.d.). European Convention on Human Rights (1950). *Encyclopedia of Public International Law, Ed. by Rudolf Bernhardt,* 2, 188–196.
F-Secure. (2018, May 25). *The Big Idea Behind GDPR – Safe and Savvy Blog by F-Secure.* Retrieved from

https://safeandsavvy.f-secure.com/2018/05/25/the-big-idea-behind-gdpr/
Gibbons, N. (2017, April 19). *EU GDPR – Organisations can run but they cannot hide*. Retrieved from https://nrgfxit.net/2017/04/19/eu-gdpr-organisations-can-run-but-they-cannot-hide/
GIGYA. (2017). Survey Report: *How Consumers Feel About Data Privacy in 2017*. Retrieved from https://www.gigya.com/resource/report/2017-state-of-consumer-privacy-trust/
Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Piscataway: Transaction.
Greenfield, P. (2018, March 25). The Cambridge Analytica files: the story so far. *The Guardian*. Retrieved from http://www.theguardian.com/news/2018/mar/26/the-cambridge-analytica-files-the-story-so-far
Grogan, S., & McDonald, A. M. (2016). Access Denied! Contrasting Data Access in the United States and Ireland. *Proceedings on Privacy Enhancing Technologies, 2016*(3), 191–211.
Haren, J. (2017, November 8). *A short history of data protection in Europe*. Retrieved from https://www.linkedin.com/pulse/short-history-data-protection-europe-dummies-john-haren/
Harford, T. (2014). Big data: A big mistake? *Significance, 11*(5), 14–19.
Hatmaker, T. (2018, May 25). US news sites are ghosting European readers on GDPR deadline. *TechCrunch.* Retrieved from http://social.techcrunch.com/2018/05/25/gdpr-tronc-us-media-companies/
Hern, A. (2018, May 21). Most GDPR emails unnecessary and some illegal, say experts. *The Guardian.* Retrieved from http://www.theguardian.com/technology/2018/may/21/gdpr-emails-mostly-unnecessary-and-in-some-cases-illegal-say-experts
Herrmann, D., & Lindemann, J. (2016). Obtaining personal data and asking for erasure: Do app vendors and website owners honour your privacy rights? (2nd ed.) *Computers and Society.* Retrieved from http://arxiv.org/abs/1602.01804
Hill, R. (2018, August 24). Chap asks Facebook for data on his web activity, Facebook says no, now watchdog's on the case. *The Register.* Retrieved from https://www.theregister.co.uk/2018/08/24/irish_data_protection_commish_opens_inquiry_on_facebook_data_transparency/
Hoepman, J.-H. (2009). Revocable privacy. *ENISA Quarterly Review, 5*(2), 16–17.
Holvast, J. (2008). *The Future of Identity in the Information Society*. Berlin: Springer.
Hornung, G., & Schnabel, C. (2009). Data protection in Germany I: The population census decision and the right to informational self-determination. *Computer Law & Security Review, 25*(1), 84–88.
How the GDPR impacts and suffocates small and medium businesses. (n.d.). Retrieved from https://www.i-scoop.eu/gdpr/gdpr-small-medium-businesses/
Hugo Monteiro, M. G. (2018, May 22). *The Myths of the "GDPR Apocalypse*.Retrieved from https://www.linkedin.com/pulse/myths-gdpr-apocalypse-hugo-monteiro/
Ianelli, C. J. (2018). Biobanks: let's share specimens. *Drug Discovery Today.* [Editoral]. https://doi.org/10.1016/j.drudis.2018.08.008
Inglehart, R. F. (2008). Changing Values among Western Publics from 1970 to 2006. *West European Politics, 31*(1-2), 130–146.
Ruiz, J., & Johnson-Williams, E. (2018). *Debates, awareness, and projects about GDPR and data protection*. Interim Report for the Information Commissioner's Office for the project: "Making new privacy rights protect and enable people's financial futures." Retrieved from https://www.openrightsgroup.org/about/reports/debates-awareness-and-projects-about-gdpr-and-data-protection
Inventis. (n.d.). *HANDLEIDING Inzage- en verwijderverzoeken Voor implementatie van de Algemene verordening gegevensbescherming (Avg*). Retrieved from https://inventus.online/.../20180202_Handleiding-inzage-en-verwijderverzoeken.doc
Jones, D. (2018, June 4). *Don't forget SARs in your GDPR content strategy.* Retrieved from https://www.itproportal.com/features/dont-forget-sars-in-your-gdpr-content-strategy/
Jones, H., & Soltren, J. H. (2005). Facebook: Threats to privacy. *Project MAC: MIT Project on Mathematics and Computing, 1*, 1–76.
Joseph, S. (2018, June 25). A month after GDPR takes effect, programmatic ad spend has started to recover.

Digiday. Retrieved from https://digiday.com/marketing/month-gdpr-takes-effect-programmatic-ad-spend-started-recover/

Kambhatla, N., & Leen, T. K. (1997). Dimension Reduction by Local Principal Component Analysis. *Neural Computation, 9*(7), 1493–1516.

Karbaliotis, C. (2017, March 9). *The Nightmare Letter: A Subject Access Request under GDPR.* Retrieved from https://www.linkedin.com/pulse/nightmare-letter-subject-access-request-under-gdpr-karbaliotis/

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1(*1), 1-12.

Koops, B.-J. (2014). The trouble with European data protection law. *International Data Privacy Law, 4*(4), 250–261.

Kottasová, I. (2018, May 11). *These companies are getting killed by GDPR.* Retrieved from https://money.cnn.com/2018/05/11/technology/gdpr-tech-companies-losers/index.html

Krigsman, M. (2015, May 30). Harvard medical professor: Big data and analytics help cure cancer. *ZDNet.* Retrieved from http://www.zdnet.com/article/harvard-medical-professor-big-data-and-analytics-help-cure-cancer/

L'Hoiry, X. D., & Norris, C. (2015). The honest data protection officer's guide to enable citizens to exercise their subject access rights: lessons from a ten-country European study . *International Data Privacy Law, 5*(3), 190–204.

Lalla, M. (2017). Fundamental characteristics and statistical analysis of ordinal variables: a review. *Quality & Quantity, 51*(1), 435–458.

Lanxon, N., & Bodoni, S. (2018, July 27). Facebook, Twitter Say Europe's Privacy Law Causing User Drop. *Bloomberg News.* Retrieved from https://www.bloomberg.com/news/articles/2018-07-27/facebook-says-eu-privacy-law-caused-user-drop-europe-disagrees

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change. *Policy Sciences, 45*(2), 123–152.

Lukács, A. (n.d.). *What is privacy? The history and definition of privacy.* Retrieved from http://publicatio.bibl.u-szeged.hu/10794/7/3188699.pdf

Machanavajjhala, A., Kifer, D., Abowd, J., Gehrke, J., & Vilhuber, L. (2008*). Privacy: Theory Meets Practice on the Map, IEEE 24th International Conference on Data Engineering, Cancun, April 4 2008 to April 12 2008*. Washington, DC, USA: IEEE Computer Society.

Mahieu, R., Asghari, H., & van Eeten, M. (2017, December). *Collectively Exercising the Right of Access: Individual Effort, Societal Effect*. Paper presented at Giganet Annual Symposium, Geneva, Switzerland.

Mansfield-Devine, S. (2015). The privacy dilemma. *Network Security, 2015*(2), 5–10.

Marr, B. (2015, June 2). Big Data At Dickey's Barbecue Pit: How Analytics Drives Restaurant Performance*. Forbes Magazine*. Retrieved from https://www.forbes.com/sites/bernardmarr/2015/06/02/big-data-at-dickeys-barbecue-pit-how-analytics-drives-restaurant-performance/

McNamee, J. (2015). *Activist guide to the Brussels maze.* Brussels: EDR. Retrieved from https://edri.org/files/activist_guide_to_the_EU_2012.pdf

Martijn, M., & Tokmetzis, D. (2016). *Je hebt wél iets te verbergen: over het levensbelang van privacy.* Amsterdam: de Correspondent.

McMullan, T. (2018, May 25). *From Razer to Yeelight, Instapaper and Unroll.me: More services go dark across the UK as companies miss GDPR deadline*. Retrieved from http://alphr.com/go/1009421

Michael, M. G. & Michael, K. (2014) *Uberveillance and the social implications of microchip implants : emerging technologies*. Hershey, PA : Information Science Reference.

Morgan, J. (2014, August 19). Privacy Is Completely And Utterly Dead, And We Killed It. *Forbes Magazine*. Retrieved from https://www.forbes.com/sites/jacobmorgan/2014/08/19/privacy-is-completely-and-utterly-dead-and-we-killed-it/

Morse, J. M. (1994). Designing funded qualitative research. In Denizin, N. K. & Lincoln, Y. S., *Handbook of qualitative research (2nd Ed*). Thousand Oaks, CA: Sage Publications.

Morsink, J. (1999). The Universal Declaration of Human Rights: Origins, Drafting, and Intent. *Pennsylvania: University of Pennsylvania Press.*

Murphy, H. (2018, July 2). Companies under strain from GDPR requests. *Financial Times.* Retrieved from

https://www.ft.com/content/31d9286a-7bac-11e8-8e67-1e1a0846c475
My Data Done Right. (n.d.). Retrieved from https://mydatadoneright.eu/
Naik, K., & Joshi, A. (2017). *Role of Big Data in various sectors.* Paper presented at International Conference on I-SMAC, Coimbatore, India.
National Academy of Engineering. (n.d.). *Grand Challenges - Secure Cyberspace.* Retrieved from http://www.engineeringchallenges.org/challenges/cyberspace.aspx
Norris, C., de Hert, P., L'Hoiry, X., & Galetta, A. (2017). *The Unaccountable State of Surveillance: Exercising Access Rights in Europe*. Berlin, Germany: Springer.
Noyb.eu. (n.d.) *My Privacy is none of your business*. (n.d.). Retrieved from https://noyb.eu/
Orwell, G. (1949). *Nineteen eighty-four*. London: Secker and Warburg.
Overkleeft-Verburg (n.d.) *Grondwet, artikel 10 (1815, august 24*). Retrieved from https://www.nederlandrechtsstaat.nl/grondwet/artikel.html?artikel=10
Ovey, C., & White, R. C. A. (2006*). Jacobs And White: The European Convention on Human Rights*. Oxford: Oxford University Press.
Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing, 19*(1), 27–41.
Pokharel, S. (2013). The Rise of Big Data: How It's Changing the Way We Think. *CFA Digest, 43*(4).
Ponemon institute. (2018). *The Race to GDPR: A Study of Companies in the United States & Europe.* Retrieved from https://iapp.org/media/pdf/resource_center/Ponemon_race-to-gdpr.pdf
Press Association. (n.d.). Data protection complaints nearly double in three months since GDPR – regulator. *Daily Gazette.* Retrieved from http://www.gazette-news.co.uk/news/national/16598430.data-protection-complaints-nearly-double-in-three-months-since-gdpr-regulator/
Privacy Europe. (2018). *European privacy framework - PRIVACY EUROPE*. Retrieved from https://www.privacy-europe.com/european-privacy-framework.html
Reding, V. (2010). A comprehensive approach on personal data protection in the European Union. *European Commission.* Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC0609&from=EN
Reding, V. (2011). The upcoming data protection reform for the European Union. *International Data Privacy Law, 1*(1), 3–5.
Regulation (EU) 2016/679. (2016). 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). Retrieved from http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679
Right of access. (2018). Retrieved from https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/
Rijo, D. (2018, May 24). *NPR & GDPR: Users that decline cookies sent to a plain text website.* Retrieved from https://ppc.land/npr-gdpr-users-that-decline-cookies-sent-to-a-plain-text-website/
Robinson, N., Graux, H., Botterman, M., & Valeri, L. (2009). Review of the European data protection directive. *Rand Europe.* Retrieved from https://www.researchgate.net/profile/Lorenzo_Valeri2/publication/265450064_Review_of_the_European_Data_Protection_Directive/links/54a93fd90cf2eecc56e69263/Review-of-the-European-Data-Protection-Directive.pdf
Roosevelt, F. (1941). *Annual address to congress--the "four freedoms."* Franklin D. Roosevelt. Presidential Library and Museum.
Rouvroy, A., & Poullet, Y. (2009). The Right to Informational Self-Determination and the Value of Self-Development: Reassessing the Importance of Privacy for Democracy. *Reinventing Data Protection?, 45–76.*
Rudgard, S. (n.d.). *Origins and historical context of data protection law.* Retrieved from https://iapp.org/media/pdf/publications/European_Privacy_Chapter_One.pdf
SAS. (2015). *Finding the Right Balance Between Personalization and Privacy. SAS.* Retrieved from

https://www.sas.com/content/dam/SAS/en_us/doc/research1/balance-between-personalization-privacy-107399.pdf
Schoneveld, D., Spanninga, H., Sprenger, J., Postma, R. (2017). *Onderzoek inzage persoonlijke gegevens* (Research report by Berenschot). Retreived from https://www.rijksoverheid.nl/documenten/rapporten/2017/08/29/onderzoek-inzage-persoonlijke-gegevens
Schiff, A. (2018, April 18). Verve Closes European Business Thanks To GDPR. *AdExchanger.* Retrieved from https://adexchanger.com/mobile/verve-closes-european-business-thanks-to-gdpr/
Scott, M., & Cerulus, L. (2018, January 31). *Europe's new data protection rules export privacy standards worldwide.* Retrieved from https://www.politico.eu/article/europe-data-protection-privacy-standards-gdpr-general-protection-data-regulation/
Sehested, T. (2018, July 10). Compliance Can Make Or Break Your organziation's Reputation. *Forbes Magazine*. Retrieved from https://www.forbes.com/sites/forbestechcouncil/2018/07/10/compliance-can-make-or-break-your-companys-reputation/
Seiler, D. (2016, January 3). General Data Protection Regulation – The end of a paper tiger [Blog]. Retrieved from https://blog.kpmg.ch/general-data-protection-regulation-the-end-of-a-paper-tiger/
Sesam (n.d.). *GDPR data access portal.* Retrieved from https://docs.sesam.io/gdpr-data-access-portal.html
Shepherd, A. (n.d.). *Firms might buy, not build, their way to GDPR compliance.* Retrieved from http://www.itpro.co.uk/data-protection/28627/firms-might-buy-not-build-their-way-to-gdpr-compliance
Singer, N. (2018, May 27). The Next Privacy Battle in Europe Is Over This New Law. *The New York Times*. Retrieved from https://www.nytimes.com/2018/05/27/technology/europe-eprivacy-regulation-battle.html
Skousen, M. (2002, May 1). *The Right to Be Left Alone*. Retrieved from https://fee.org/articles/the-right-to-be-left-alone/
Sonos (2018, April 24). We're Updating the Sonos Privacy Statement [Blog]. Retrieved August 25, 2018, from https://blog.sonos.com/en-gb/updating-sonos-privacy-statement/
Spiller, K. (2016). Experiences of accessing CCTV data: The urban topologies of subject access requests. *Urban Studies , 53*(13), 2885–2900.
State of Callifornia. (2018, June 29). Assembly Bill No. 375. *California Legislative Information.* Retrieved from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375
Statt, N. (2018, August 7). *Instapaper returns for EU users post-GDPR with new premium subscription option*. Retrieved from https://www.theverge.com/2018/8/7/17660420/instapaper-gdpr-user-privacy-eu-return-premium-relaunch
Sturges, P. (2005). Is privacy dead? *CILIP Update, 4*(11), 16-18.
Sunstein, C. R. (2003). *Why Societies Need Dissent*. Harvard: Harvard University Press.
Survey: 61 percent of companies have not started GDPR implementation. (n.d.). Retrieved from https://iapp.org/news/a/survey-61-percent-of-companies-have-not-started-gdpr-implementation/
Swire, P. P., Litan, R. E., & Litan, R. E. (1998). None of your business: world data flows, electronic commerce, and the European privacy directive. *Harvard Journal of Law & Technology, 12*(3), 268-288.
The Economist. (2017). *Data is giving rise to a new economy*. Retrieved from https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy
The EU GDPR. (n.d.). Retrieved from https://www.paconsulting.com/insights/2017/eu-general-data-protection-regulation/
The powers of attorney. (2016, March 23). Retrieved from https://diplomatie.belgium.be/en/services/services_abroad/notary_expertise/powers_of_attorney
Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review, 34*(1), 134–153.
Tracey, J. M., Arroll, B., Richmond, D. E., & Barham, P. M. (1997). The validity of general practitioners' self-assessment of knowledge: cross sectional study. *BMJ , 315*(7120), 1426–1428.
Trendy Dots (2018, May 27). Even tho the topic of #GDPR is widely discussed and covered in countless

internet posts, the retention rate of the concept in followup interest, manifested in Google searches is mostly focused on the territory of the #EuropeanUnion [Tweet]. Retrieved from https://twitter.com/trendydots/status/1000740185037049856

Trevor Hughes, J. (2016, January 11). General Data Protection Regulation: A Milestone Of The Digital Age. *TechCrunch.* Retrieved from http://social.techcrunch.com/2016/01/10/the-biggest-privacy-law-in-the-world-has-arrived/

TrustArc. (n.d.). *Individual Rights Manager*. Retrieved from https://www.trustarc.com/products/individual-rights-manager/

Tuite, K., Snavely, N., Hsiao, D.-Y., Tabing, N., & Popovic, Z. (2011). *PhotoCity: Training Experts at Large-scale Image Acquisition Through a Competitive Game.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, Canada.

U.S. chamber (2015) *U.S. Chamber Statement on EU General Data Protection Regulatio*n Retrieved from https://www.uschamber.com/press-release/us-chamber-statement-eu-general-data-protection-regulation

United Nations. (2017). Sustainable Development Goals. In *Integrated Approaches for Sustainable Development Goals Planning: The Case of Sustainable Development Goal 6 on Water and Sanitation,* New York: United Nations.

Urdan, T. C. (2016). Statistics in Plain English, Fourth Edition. *Abingdon: Taylor & Francis.*

Urglavitch, M. (2017, October 17). *OneTrust Launches First-to-Market Data Subject Access Request (DSAR) Portal to Simplify GDPR*. Retrieved from https://www.onetrust.com/onetrust-launches-first-market-data-subject-access-request-dsar-portal-simplify-gdpr-compliance/

Ustaran, E. (2017, September 22). GDPR – beyond the panic - Internet Newsletter for Lawyers. *Onetrust.* Retrieved from http://www.infolaw.co.uk/newsletter/2017/09/gdpr-beyond-panic/

Verhagen, L. (2018, August 10). Klanten hebben recht op hun persoonlijke gegevens - maar krijgen ze die ook? We testten bekende bedrijven. *Volkskrant.* Retrieved from https://www.volkskrant.nl/gs-b7e6d7c8

Voorbeeldbrief verzoek om inzage.pdf. (n.d.). Retrieved from https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/voorbeeldbrief_verzoek_om_inzage.pdf

Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review, 4*(5), 193–220.

Waxman, O. B. (2018, May 24). The GDPR Is Just the Latest Example of Europe's Caution on Privacy Rights. That Outlook Has a Disturbing History. *Time*. Retrieved from http://time.com/5290043/nazi-history-eu-data-privacy-gdpr/

Westin, A. (1970). Privacy and freedom. *London: The Bodley Head.*

Westin, A. F. (2003). Social and Political Dimensions of Privacy. *The Journal of Social Issues, 59*(2), 431–453.

Wilhelm, E. (n.d.). *A brief history of the General Data Protection Regulation.* Retrieved from https://iapp.org/resources/article/a-brief-history-of-the-general-data-protection-regulation/

Winter, H. B., De Jong, P. O., Sibma, A., Visser, F. W., Herweijer, M., Klingenberg, A. M., & Prakken, H. (2008). *Wat niet weet, wat niet deert: een evaluatieonderzoek naar de werking van de Wet bescherming persoonsgegevens in de praktijk.* Retrieved from http://repository.tudelft.nl/assets/uuid:225def96-5c92-4b86-b247-2abc52dacd47/1382b-samenvatting_tcm44-165375.pdf

Yamashita, T., Yamashita, K. & Kamimura R. (2007) *A Stepwise AIC Method for Variable Selection in Linear Regression, Communications in Statistics - Theory and Methods*, 36:13, 2395-2403, DOI: 10.1080/03610920701215639

Yu, J. (hyunjae), & Cude, B. (2009). "Hello, Mrs. Sarah Jones! We recommend this product!" Consumers' perceptions about personalized advertising: comparisons across advertisements delivered via three different types of media. *International Journal of Consumer Studies, 33*(4), 503–514.

Zwenne, G.-J., Duthler, A.-W., Groothuis, M., Kielman, H., Koelewijn, W., & Mommers, L. (2007). *Eerste fase evaluatie Wet bescherming persoonsgegevens: Literatuuronderzoek en knelpuntenanalyse.* Retrieved from https://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid%3A8a988e84-cae7-4442-9685-7e657aac74ba

Annexes

A.1 Overview of changes to GDPR

GDPR	Topic	Changes	DPD
1-11	General provisions and principles	• Increased territorial scope • New definitions for pseudonymization, binding corporate rules, data breach and health related data • New provisions & principles: transparency, accountability & processing w/out identification • Clarified principles: Data minimization, consent conditions, lawful processing • New conditions: Child's consent by parent or custodian	2, 4, 6,7, 8
12-22	Personal rights	See table 2	10-15
24-31	General obligations	• New obligation for controllers: data protection by design and by default • Clarification on controller's responsibility with multiple controllers • New controller obligation: & processers: maintain records of activity & cooperate with supervisor	16-18
32-34	Security of personal data	• Extended obligations of data security to also cover processors • New controller obligation: Notification of data breach to authority and subject • New processor obligation: Notification of data breach to controller	17, 4
35-36	DPIA and prior consultation	• New controller obligation: DPIA prior to likely risky operations • Simplified controller obligation: Prior consultation with DPA only if DPIA shows high risk	20
37-39	DPO	• New controller obligation: Requirements for DPO obligation and their core tasks	18
40-43	Codes of conduct and certification	• Implication of approval of code of conduct by DPA • New means to demonstrate compliance with GDPR, certification, seals and marks	27
44-49	International data transfer	• New conditions for personal data transfer and approval mechanisms	25-26
50-76	Supervisory authority	• Specified powers and tasks for DPO's, also in working with multinational companies • Creation of EDPB to manage disputes, give advice and guidance	
77-84	Remedies, liability and penalties	• Specification of data subjects right to complain to supervisor, or others on their behalf • Extended processor liability for damage caused to subject when infringing GDPR • Clarified liability of joint controllers & processors: each liable for entire damage • Administrative fines that supervisors are entitled to impose	22-24

GDPR	Topic	Changes	DPD
12	Transparency & modalities	• New controller obligation: transparent, concise and understandable information notices on processing	12
13-15	Information and access to personal data	See table	10-12a
16-20	Rectification and erasure	• Specified right to rectification, erasure and restriction of processing • Introduced right to be forgotten & right to data portability	12b-c
21-22	Right to object and automated decision making	• Broadened right to object, with obligation of controller to notify individuals of the right and proof it's necessity when asked. Objection to direct marketing is specifically strict. • Automated decision making is only possible with suitable measures and protection • Profiling based on sensitive data needs explicit consent, or authorization by law	14-15

Sources: (Faradina, 2017; Ausloos & Dewitte, 2018; Birds, 2017; Directive 95/46/EC, 1995; Regulation (EU) 2016/679, 2016; Tikkinen-Piri, Rohunen, & Markkula, 2018)

A.2 HREC application

Research Ethics Application

Please fill in the checklist first if you have not done so already. Please complete this form digitally and send it the Ethics Committee.

Date of Submission:6-1-2018

Project Title: Access rights in the GDPR era

Name(s) of researcher(s): Thomas van Biemen

Name of supervisor (if applicable): Hadi Asghari

Contact Information

Department: TPM, MAS, OG

Telephone number: +31 15 27 83433

E-mail address: h.asghari@tudelft.nl

Contact information of external partners (if applicable): n.a.

Summary

Please provide a brief summary of the research.

In the proposed research, we want to test the GDPR's subjects access rights in practice. By exercising this right, everybody should be able to request a copy of the personal data that is stored by corporations. In contrast to earlier research on older implementations of access rights, we will not only request personal data from the researcher itself but also request data on behalf of volunteers (10 per volunteer, 50 volunteers) This collaborative effort is meant to provide us with a bigger set of responses, which makes further investigation into the differences among responses to data access requests possible.
When meeting volunteers for the first time, we will explain the procedure, go over the informed consent form and brainstorm together for possible agencies and companies that have collected the volunteers' data. Researchers will communicate with these corporations on behalf of the volunteer (hereafter: participant). The participants can choose if responses to the requests can be seen directly by the researcher or only after the participant has shared these. This choice is entirely up to the participant and can be changed per corporation at any time. The datarights.me website that has been approved and used in a previously reviewed study will be used to inform and update respondents during the process.

Research

R.1. What is the research question? Please indicate what scientific contributions you expect from the research.

How well are companies complying with the right of access regulation under the GDPR, and what can be improved to further facilitate this compliance? What are the differences between responses to access requests

under the GDPR and wbp? What differences can be found in compliance between different companies and can differences in company characteristics help predict these differences?
By answering these questions, we hope to measure the effect that introducing the GDPR regulation has on practicing the fundamental right of privacy in practice. Does stronger EU regulation really help to protect the rights of its inhabitants? When combined with future results, our research can also facilitate an investigation into compliance with the law on the longer term: does it increase or decrease? Furthermore, the novel research method and application that are used in this research will be documented so that these technological artefacts can contribute to new research. Especially the application of a research method which is unique in the data privacy field: requesting data on behalf of others, is expected to help research in this and other areas where data collection by researchers is not enough. Expected insights will be of value to academics, policy-makers, industry-thinkers, and the general public.

R.2. What will the research conducted be a part of?
☐Bachelor's thesis
☒Master's thesis
☐PhD thesis
☐Research shills training
Other, namely: Enter what the research is part of here.

R.3. What type of research is involved?
☒Questionnaire
☒Observation
☐Experiment
Other, namely: Enter the type of research here.

R.4. Where will the research be conducted?
☒Online
☐At the university
☐Off-campus / non-university setting: Enter which setting here.
Other: Access requests will be sent online or via letter on behalf of participants. Communication is stored securely on TU Delft servers.

R.5. On what type of variable is the research based?
Give a general indication, such a questionnaire scores, performance on tasks, etc.
The research will be based on a questionnaire score on how well responses to data requests relate to how responses should be under new regulation (and how these response scores relate to those of other companies). It will thus be based on a questionnaire score to grade observations.

R.6. If the research is experimental, what is the nature of the experimental manipulation?
It is possible that the communication affects the organization's practices (e.g. awareness on how to handle access requests) and the participant's psyche (e.g. after learning what is known about them). We expect these affects mostly positive, i.e., increased awareness, but outside of our research scope.

R.7. Why is the research socially important? What benefits may result from the study?
This research will show how well the right of access is implemented in practice. This right is seen as the "natural precondition" to guaranteeing privacy in a modern world. The GDPR provides users with stronger tools to decide on their own balance between the risks and benefits of data collection and analysis. Our research investigates how this new regulation works out in practice. We will thus investigate a possible solution to fundamental questions on assure fundamental human rights trough regulation. From a bigger perspective, results can help substantiate the discussion whether (European) laws in general can help us to balance these fundamental rights.

R.8. Are any external partners involved in the experiment? If so, please name them and describe the way they are involved in the experiment.

Platform developer(s): the online platform that is used in the research is going to be built by an external programmer and designer. Hadi will keep close contact with these partners to guarantee the necessary security and privacy.

Participants

Pa.1. What is the number of participants needed? Please specify a minimum and maximum.
Minimum: 10
Maximum: 100

We want to send requests to 10 companies on behalf of each participant, creating a dataset of 100-1000 responses. (expected participants = 50)

Pa.2.a. Does the study involve participants who are particularly vulnerable or unable to give informed consent? *(e.g., children, people with learning difficulties, patients, people receiving counselling, people living in care or nursing homes, people recruited through self-help groups)*
No.

Pa.2.b. If yes and unable to give informed consent, has permission been received from caretakers/parents?
Non-applicable

Pa.3. Will the participants (or legal guardian) give written permission for the research with an 'Informed Consent' form that states the nature of the research, its duration, the risk, and any difficulties involved? If no, please explain.
Yes, and a researcher will be with the participant to answer any further questions that may arise.

Pa.4. Are the participants, outside the context of the research, in a dependent or subordinate position to the investigator (such as own children or students)? If yes, please explain.
no

Pa.5. How much time in total (maximum) will a participant have to spend on the activities of the study?
Everybody will spend 2 times 1 hour for intake and wrap-up talks. Further time investments are up to the participants. If the participants wants to read all communication that is sent on their behalf, it will take at most two hours per company. The maximum will thus be 22 hours, although I expect that no participant will do this.

Pa.6. Will the participants have to take part in multiple sessions? Please specify how many and how long each session will take.
Yes, 1 kick-off session of 1 hour and 1 wrap-up session of 1 hour.

Pa.7. What will the participants be asked to do?
Kick off session: To read and sign the informed consent form and to brainstorm for possible companies to contact with the researcher. The participant will also be asked to choose if responses from these companies can be shared with the researchers directly or only after reviewing the content (this can be changed per company at any time in the online environment) In between: Read responses and decide of these can be shared with the researcher. Review the final responses for compliance with the law if data is not shared with the researcher. Wrap-up session: General talk on how to handle communication with companies that are still not compliant after the research (e.g. do you want to submit a complaint to the authority?) and on how the innovative research set-up was perceived by participants. This session also serves as a deadline for any data that still has to be classified.

Pa.8. Will participants be instructed to act differently than normal or be subject to certain actions which are not normal? *(e.g. subject to stress inducing methods)*
no

Pa.9. What are the possible (reasonably foreseeable) risks for the participants? Please list the possible harms if any.
Two potential risks are expected and managed by asking the participant to choose between two modes of communication, these are: (1) Risk to the privacy of the volunteers. Responses to data access requests almost always have a risk to contain personal information. Therefore, participants can choose to review the responses before these are shared with the researchers. It is also possible for the participant to participate in the research without ever sharing the responses with the researcher: participants will then be asked to classify the response themselves. Only the classification will be shared with the researcher. (2) Risk to over-burden the volunteers. Data access requests can be very burdensome as companies are not expected to respond with all requested information after the first communication period. Therefore, communication with companies will be performed by the researchers on behalf of the participants. Participants can also choose to share the responses automatically with the researcher to streamline the back-and-forth communication process. Of course, all communication with the company will still be shared with the participant. Participants will be asked to choose either the strict or lenient sharing rules for responses immediately after brainstorming for possible companies and will be able to change these settings (or stop the communication altogether) per company at any time through the datarights.me website.
This research deals with personal data. We took extra measures to protect their identity and data. Privacy is the key features of the system design. The collected data will be hosted at TU Delft, encrypted in transit and at rest.

Pa.10. Will extra precautions be taken to protect the participants? If yes, please explain.
To guarantee the precautions that are described in the previous answer (allowing the participant to not share any or communication or received data with the researchers if they choose so), the datarights.me system is designed in a way that this is also not physically possible for the researchers. Researchers will only be able to see the subject lines of responses to sort responses with the right participant & company if the sorting algorithm cannot do this automatically.

Pa.11. Are there any positive consequences for a participant by taking part in the research? If yes, please explain.
Yes, the participants can send out data requests without having to partake in the long communication procedure that often accompanies this process. Participants have already indicated that they are very curious about responses from certain companies. In the end, we hope that participants feel more empowered and are more aware of their privacy rights.

Pa.12. Will the participants (or their parents/primary caretakers) be fully informed about the nature of the study? If no, please explain why and state if they will receive all information after participating.
yes

Pa.13. Will it be made clear to the participants that they can withdraw their cooperation at any time?
Yes, this is also added to the consent form

Pa.14. Where can participants go with their questions about the research and how are they notified of this?
To the researcher (Thomas). All participants have the contact information of this researcher.

Pa.15. Will the participants receive a reward?
☐Travel expenses
☐Compensation per hour
☒Nothing
Other, namely: Enter the reward here.

Pa.16. How will participants be recruited?
Volunteers will be selected by the researchers from his own network and asked to participate.

Privacy

Pr.1. Are the research data made anonymous? If no, please explain.
Yes, the data will be anonymized as soon as possible. After receiving personal data from companies, the participant will be able to choose to either classify the data themselves, (partly) share the data with the researcher to classify or delete altogether. After the data is classified (either by the participant or the researcher), the data itself will be anonymized. Only anonymous data of the responses will thus be stored for the remainder of the research. When data collection and classification has stopped, and the researcher has finished the final meeting with the respondent, the classification itself will also be anonymized.

Pr.2. Will directly identifiable data (such as name, address, telephone number, and so on) be kept longer than 6 months? If yes, will the participants give written permission to store their information for longer than 6 months?
No, this data will not be stored longer than 6 months. For transparency sakes, we still ask for respondents written permission to store the data until the end of the research, which is expected to be in October 2018.

Pr.3. Who will have access to the data which will be collected?
Thomas van Biemen will have access to all data that respondents are willing to share with researchers. His personal computer will also contain information on the distribution of companies and participants, and the basic and contact information of volunteers. We will only save the information that is vital to contact the participants and report on possible biases within the population (e.g. mostly under 35, mostly students, or very privacy aware in general). This data will be saved anonymously, and password protected until data analysis is completed. The key to link this data with requests and responses data (vital for data analysis to check for possible biases) will be saved in a separate file and protected with a different password.

Hadi Asghari will also have access to all data that respondents are willing to share with researchers. The respondents informed consent form explicitly names Hadi and Thomas and their ability to see this data.

Martijn Marnier, the chairman of the thesis committee, will receive updates on the data collection progress, but these will be anonymous.

Respondents will have access to all data that we received from them, all communication that happens on their behalf, their own personal data and how it is classified. Respondents will also always know if they have shared any of this information with the researchers.

Pr.4. Will the participants have access to their own data? If no, please explain.
Yes, please see previous answer.

Pr.5. Will covert methods be used? *(e.g. participants are filmed without them knowing)*
No, Participants will be fully aware of the data methods used.

Pr.6. Will any human tissue and/or biological samples be collected? *(e.g. urine)*
no

Documents

Please attach the following documents to the application:

- Text used for ads (to find participants); not applicable
- Text used for debriefings; not applicable, this will be done face to face.
- Form of informed consent for participants; added
- Form of consent for other agencies when the research is conducted at a location (such as a hospital or school). Not applicable

A.3 Informed consent form

The following consent form was used to structure the kick-off meeting and systemically request the required information. It is loosely based on the example that was provided by the Delft University of Technology HREC team around March 2018. All volunteers that participated in the research have read and filled out this document.

Exercising access rights in the GDPR era

Informatie voor deelnemers

Inhoud:

Achtergrond en rolverdeling

Om eerdere Europese richtlijnen te harmoniseren en beveiliging van persoonlijke data te versterken is in de hele EU op 25 mei 2018 de algemene verordening gegevensbescherming (AVG, GDPR) in werking gegaan. Een belangrijk onderdeel van deze wetgeving zijn de zogenaamde ARCO-rechten van betrokkenen. Dit is het recht op toegang (access), correctie (rectification), annulering (cancellation) en verzet (opposition) tot/tegen verwerking van persoonlijke data door betrokkenen. Het recht op toegang tot de verzamelde data is hierin een belangrijke eerste stap. Betrokkenen kunnen immers lastig een correctie van hun data aanvragen wanneer de data zelf niet bekend is.

Uit eerder onderzoek blijkt dat aanvragen voor datatoegang in het verleden door veel bedrijven en instanties niet volgens de wet werden afgehandeld. Zo werd vaak te laat, verkeerd of niet gereageerd op aanvragen. Ondanks een aantal voorbeelden van instanties die er correct en positief mee omgaan lijkt het overgrote deel van de Nederlandse bedrijven en instanties het recht op toegang niet te kennen, niet te kunnen of niet te willen honoreren. Dit ondanks het feit dat dit recht sinds en 1995 in de Europese en sinds 2001 in de Nederlandse wetgeving is opgenomen.

In dit onderzoek zal worden gekeken naar de naleving van het recht op data-toegang voor betrokkenen in Nederland na invoering van de AVG/GDPR. Hiermee zal worden onderzocht of een strengere wet die van bovenaf wordt opgelegd, en de hernieuwde aandacht en discussie over privacy die hieraan parallel loopt, in de praktijk hebben bijgedragen aan een betere naleving van het recht op datatoegang. Daarnaast zal worden onderzocht of verwachtte verschillen in reacties van bedrijven verklaard kunnen worden door te kijken naar bijvoorbeeld de leeftijd, grootte of beroepstak van dat bedrijf.

Waar eerder onderzoek slechts de data-opvragen van de onderzoeker zelf omvatten, is voor dit onderzoek een grotere opzet nodig. Door namens meerdere personen data op te vragen kan een dataset worden verzameld met een groter aantal unieke bedrijven, wat resultaten zekerder maakt. Om niet te veel van deelnemers te vragen, zal de communicatie met bedrijven altijd door de onderzoeker worden gedaan. Deze communicatie kan door de deelnemer online gevolgd worden.

Wanneer een instantie data opstuurt, is het altijd aan de overeenkomende deelnemer wat ermee gebeurt. Verderop in dit document kun je zelf invullen bij welke instantie namens jou data opgevraagd mag worden. Per instantie kun je vervolgens aangeven hoe met de opgevraagde data omgegaan dient te worden, je kunt kiezen uit de volgende opties:

- Vertrouwelijk. Teruggestuurde data zal alleen te zien zijn voor de deelnemer zelf. De deelnemer heeft na het inzien van de data de keuze om deze zelf te classificeren of om dit (deels) toch door de onderzoeker te laten doen.
- Vrij. Teruggestuurde data is zowel voor de deelnemer als de onderzoeker te zien. Het classificeren van de data kan hierdoor geheel door de onderzoeker en zonder tijdsinvestering van de deelnemer gebeuren.

Op een online dataprivacy-portaal kan de deelnemer deze opties per instantie op elk moment aanpassen of terugtrekken. De rapportage die na dit onderzoek volgt zal alleen geaggregeerde en geanonimiseerde data bevatten. De geanonimiseerde data zal gedurende een periode van vijf jaar worden opgeslagen door de TU Delft en kan gebruikt worden in verder onderzoek (bijvoorbeeld om de conclusies van dit onderzoek te controleren). Verder verkregen data, waaronder de door instanties toegestuurde data en andere persoonlijke informatie zal na afsluiting van het onderzoek worden verwijderd.

Basisinformatie

Deze gegevens zullen worden gebruikt om contact te houden of gebruikt worden om resultaten te controleren op biasses.

Contactgegevens:

Naam: ……

Telnr. (niet verplicht): …….

Email: ……..

Algemene informatie:

Geboortedatum: ……..

Geslacht: …….

Beroep: ……

Nationaliteit: ……

Betrokkenheid
Hoe zou u zichzelf omschrijven:

Kennis over:	(Zeer laag)		(gemiddeld)		(zeer hoog)
ICT	0	0	0	0	0
Privacywetgeving	0	0	0	0	0

Betrokkenheid bij:	(Totaal niet)		(gemiddeld)		(Heel erg)
ICT	0	0	0	0	0
Privacy	0	0	0	0	0

Maakt u gebruik van speciale privacytoepassingen?
(bijvoorbeeld een addblocker of vpn)
0 Nee

0 Ja, namelijk:

……

……

Mogelijke bedrijven/instanties

Naam:	Data-omgang:	
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk
……………………………………………………………………………………	0 Vrij	0 Vertrouwelijk

Naam:	Data-omgang:	
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk

Naam:	Data-omgang:	
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk
……	0 Vrij	0 Vertrouwelijk

Toestemming onderzoek

Titel onderzoek: Exercising right of access in the era of GDPR
Verantwoordelijke onderzoeker: Thomas van Biemen, begeleid door Hadi Asghari

In te vullen door de deelnemer

	Ja	Nee
Ik heb de informatie op eerdere pagina's gelezen en begrepen. Verdere vragen zijn naar tevredenheid beantwoord.	0	0
Ik stem ermee in dat verkregen persoonlijke data zal worden opgeslagen wanneer dit relevant is voor het onderzoek en weet dat deze data zo snel mogelijk zal worden geanonimisserd, in ieder geval na afronden van het onderzoek (verwachte einddatum 30 oktober 2018).	0	0
Ik weet en stem ermee in dat de gegevens en resultaten van het onderzoek alleen anoniem, geaggregeerd en vertrouwelijk aan derden bekend gemaakt zullen worden als onderdeel van de rapportage van een masterscriptie.	0	0
Ik begrijp dat geanonimiseerde data binnen de TU Delft voor verdere onderzoeksdoeleinden en controle van de onderzoeksresultaten voor een periode van maximaal vijf jaar opgeslagen kunnen worden.	0	0
Ik ben ervan op de hoogte dat deelname aan dit onderzoek een risico met zich mee kan brengen voor mijn privacy en die van mijn gegevens. Ook ben ik op de hoogte van de maatregelen die zijn genomen om dit risico te minimaliseren.	0	0
Ik weet dat van mij enige inzet verwacht wordt om de communicatie met bedrijven te lezen en te classificeren wanneer ik dit niet door de onderzoeker wil laten doen.	0	0
Daarnaast ben ik ervan op de hoogte dat er aan het eind van het onderzoek een "wrap up session" wordt georganiseerd van ongeveer een uur waaraan ik verwacht wordt om deel te nemen.	0	0
Ik stem geheel vrijwillig in met deelname aan dit onderzoek. Ik behoud me daarbij het recht voor om op elk moment zonder opgaaf van redenen mijn deelname aan dit onderzoek te beëindigen.	0	0
Voor elk bedrijf waarvoor ik toestemming heb gegeven om in mijn naam data op te vragen heb ik aangegeven in welke mate ik mijn data met de onderzoeker wil delen. Ik weet dat ik ook deze keuze op elk moment en zonder opgaaf van redenen kan veranderen.	0	0
De contactgegevens van de uitvoerende onderzoeker zijn bij mij bekend. Daarnast weet ik dat ik met deze onderzoeker contact op mag nemen met vragen.	0	0

Naam deelnemer: ……..

Datum: …… Handtekening deelnemer: ……

In te vullen door de uitvoerende onderzoeker

Ik heb een mondelinge en schriftelijke toelichting gegeven op het onderzoek. Ik zal resterende vragen over het onderzoek naar vermogen beantwoorden. De deelnemer zal van een eventuele voortijdige beëindiging van deelname aan dit onderzoek geen nadelige gevolgen ondervinden.

Naam onderzoeker: ……..

Datum: …… Handtekening onderzoeker: ……

Toestemming communicatie t.b.v. aanvragen data-toegang

Geachte heer, mevrouw,

Ik, …… (hierna te noemen: deelnemer), verklaar hierbij toestemming te geven aan Thomas van Biemen en Hadi Asghari (beide hierna te noemen: uitvoerende) om:

- Namens mij inzageverzoeken op grond van de Algemene Verordening Gegevensbescherming te versturen naar instanties waarvan redelijkerwijs verwacht kan worden dat ze mijn gegevens in bezit hebben.
- Namens mij te communiceren met relevante instanties met als doel het verkrijgen van gegevens die onder de algemene verordening gegevensbescherming vallen. Communicatie via de uitvoerenden heeft hierbij voor mij altijd de voorkeur boven persoonlijke communicatie, tenzij instanties wensen om mijn identiteit met meer zekerheid vast te stellen.
- Mogelijke persoonsgegevens of andere gegevens die uit deze communicatie worden ontvangen op te slaan tot ik aangeef welke gegevens bewaard en/of met hen gedeeld mogen worden.

Daarnaast verklaar ik op de hoogte te zijn van:

- De vertrouwelijke aard van de informatie die in eerdergenoemde communicatie naar voren kan komen en de maatregelen die zijn getroffen om hier mee om te gaan. (zoals een beveiligd systeem om de gegevens op te slaan, het versleuteld communiceren van de gegevens, **…**)

…….. (naam deelnemer)	…….. (naam uitvoerende)
…….. (datum)	…….. (datum)
…….. (handtekening)	…….. (handtekening)

A.4 Standard request letter

Thomas van Biemen & Hadi Asghari
Jaffalaan 5, 2628 BX Delft
thomas.biemen@datarights.me

[Bedrijfsnaam]
[Bedrijfsadres]

Betreft: Inzageverzoek [Bedrijfsnaam] op grond van de AVG

Geachte heer, mevrouw, Delft, 21-07-2018

Namens mevrouw [Volunteer] wil ik middels deze brief gebruikmaken van het recht op inzage zoals te vinden in artikel 15 van de Algemene Verordening Gegevensbescherming (AVG, GDPR). Conform deze wet ontvang ik graag van u:

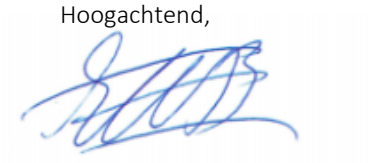
- Een overzicht van de categorieën persoonsgegevens die u van [Volunteer] verwerkt. (Artikel 15, lid 1, b)
- Een kopie van de feitelijke gegevens die u van [Volunteer] verwerkt. (Artikel 15, lid 3)
- Wat het doel is van de verwerking van deze gegevens. (Artikel 15, lid 1, a)
- Met wie deze gegevens zijn gedeeld of zullen worden gedeeld. (Artikel 15, lid 1, c)
- De periode dat deze gegevens naar verwachting zullen worden opgeslagen. (Artikel 15, lid 1, d)
- Alle mogelijke informatie over de bron van deze gegevens. (Atikel 15, lid 1, g)
- Het bestaan van automatische besluitvorming, informatie over de onderliggende logica, het belang en de verwachte gevolgen van deze automatische besluitvorming. (Artikel 15, lid 1, h)

Ik verzoek u deze data voor zover mogelijk in beveiligde vorm in een gangbaar, gestructureerd en machineleesbaar formaat, maar in ieder geval per email naar het volgende adres te versturen: [Volunteer email]. Mocht u toch besluiten om de gegevens op een andere manier te versturen, dan verzoek ik u om dit van tevoren te laten weten. Zoals ook te vinden in artikel 12, lid 3 en 4 van de AVG verwacht ik uw antwoord in ieder geval binnen het wettelijk termijn van één maand.

Als bijlage bij deze brief vindt u een ondertekende machtigingsbrief waarin [Volunteer] aangeeft dat ondergetekende namens haar communiceert. Als verder bewijs van identificatie vindt u ook een voor AVG aanvragen gebruikelijke kopie van de identiteitsbewijzen van zowel [Volunteer] als ondergetekende bijgevoegd. Mocht dit voor u niet genoeg zekerheid geven over [Volunteer] haar identiteit en/of machtiging, dan kunt u dit controleren door contact op te nemen haar persoonlijke mailadres: [Volunteer email] of per telefoon wanneer haar nummer bij u bekend is. Zoals ook door [Volunteer] is aangegeven in de machtigingsbrief wil ik u verzoeken om verdere communicatie zo veel mogelijk via [Volunteer email] te laten lopen, dit adres is doelbewust ontworpen om veilig en vertrouwelijk te communiceren en gegevens te ontvangen.

Voor meer informatie over relevante wetgeving, uitzonderingen en sancties die kunnen worden opgelegd naar aanleiding van dit verzoek, verwijs ik u graag naar informatie op de website van de Autoriteit Persoonsgegevens: https://autoriteitpersoonsgegevens.nl/.

Hoogachtend,



Namens [Volunteer]
Thomas van Biemen

A.4 Variable coding guidelines

Factor	Explanation
Subject factors	
Relationship	Relationship with company (customer, ex-customer, employee, ex-employee, (business) partner, other)
Volunteer ID	Unique ID per volunteer/data subject
Company factors	
Company name	
Establishment year	The date can often be found on wikipedia or company website, sometimes in yearly report Can be challenging b/c companies often change names or integrate w/ other companies horizontally/vertically. I choose the date from when the organization or a precursor started offering the services they offer now or started offering the service for which the request is relevant for bigger organizations
Size	Often found in an organizations yearly reports Using the Dutch Central statistics bureau definitions: - small< 50 employees - Medium< 250 - big>250 Many companies are part of larger holdings, count number of employees of the organization that responded to the request
Location	Most easily found on wikipedia NL, when companies only operate in The Netherlands NL_I, when an dutch company grew to also operate customers outside the country EU, International company operated from another EU country I, International company operated from outside EU
Sector	Using the first hierarchy of the CBS SBI method, an english version can be found here: https://www.cbs.nl/en-gb/our-services/methods/classifications/activiteiten/standard-industrial-classifications--dutch-sbi-2008-nace-and-isic--/the-structure-of-sbi-2008-version-2018 Dutch version can also be navigated/serached in several ways online: https://sbi.cbs.nl/cbs.typeermodule.typeerservicewebapi/content/angular/app/#/ If companies fall in multiple categories, only code for core business or relevant business w/ respect to request
Communication factors	
comm_start	How first request is made (as described in privacy statement or on website): Letter, e-mail, Webform
Responded	1, when a company has responded. 0, when a company has not responded
No_of_contact	How many mails, calls and letters have gone back and fort in the process in total
1st contact	Date that 1st request is sent
last contact	Date of last communication
Contact method	What kind of communication happened between initial request and data receiving was (Can be multiple): Letter, e-mail, Website/app/online tool, phone, in person

Checked	In my research this variable shows if the organization reached out to the data subject to ask if the request was indeed made on their behalf or check identity further then checking copy of ID
Reminders	How many reminders have been send, and if possible on what date these have been sent
Comm_res	How final results (=data, when data and answers to questions were seperately communicated) have been communicated: Mail, attachment of mail, Letter, registered letter, handed over personally, dowloaded (from online environment), app Can have multible entries if data is sent in multiple ways
Comm_res_add	Was this result send to: - req, requested adress - know, address known to organization (f.e. home address, non-datarights mail address.)
2factor	Was two factor authentication needed to receive results y/n, 1/0
Encripted	Was received data encrypted: yes1(key via sep. mail), yes2(key via other way), no
retr.	Other restrictions to data? (f.e. limited number of downloads, limited time available)
Response factors	
Req_more_time	Did the company (lawfully) ask for extension y/n, 1/0
Data_Sent	Was personal data sent y/n, 1/0
Data_correct	Was sent data correct & complete y/n, 1/0
Data_form	What format was the data in: - Machine readable - Machine parseable - Non machine-readable digital - Paper When multiple ways used, only code most machine readable way
Answered questions	Did the company answer the other questions (yes, no, partial) Partial = answered only some of the questions
Aswered_spec	Did the company answer questions specifically or generally ("see our privacy statement online") y/n, 1/0
Ex_q_n	Was an extra follow-up neccesary to receive answer to (specific) questions y/n, 1/1s/0, (1a when answers are received but request is for more specific answers)

A.5 Data analysis

Data analysis methods were used throughout this thesis research. Their application was especially vital in providing the numerical results that are reported in section. A link to the jupyter notebook files is given in the first section of this annex. Following sections include more detailed information concerning the Principal Component Analysis and volunteer analysis respectively.

A.5.1 Jupyter notebook file

All data analysis in this thesis research was performed using python packages in jupyter notebooks. These notebooks can be found via the following public GitHub repository:

https://github.com/tvanbiemen/Personal-privacy-in-practice

In this repository, two jupyter notebook files can be found. The first contains all data analysis concerning the numerical data analysis that was used for section 4.1. The second file contains the data analysis that was performed on the volunteer data and is referenced briefly in section 4.3.

A.5.2 Principal Component Analysis

The results that are presented in Figure 5 of section 4.1.1, already allow us to make some educated assumptions on how organizations respond to subject access requests and how the introduction of the GDPR influences these responses. However, this does not allow us to confidently answer the research questions posed in chapter 2. The division of results in 5 variables further complicates both the analysis and interpretation of effects that variables have on these results, especially when examining differences in more granular variable, such as sectors.

These 5 variables are introduced in chapter three to measure organizations response to subject access requests in more detail. In other words, the variables all present a different dimension in the measurement of a single underlying variable: how good a certain reply is. This assumption is tested by means of a Principal Component Analysis (PCA), in which an algorithm tries to find the best way to describe results by using fewer variables. PCA therefore serves to reduce the dimensions necessary to describe results. When multiple variables are measuring a part of the same result dimension, the analysis should show that:

1. The reduced dimensions still classify results similar to the input variables. Meaning that the reduction did not lead to a big loss in the explaining power of the variables.
2. All input variables have an influence on the value of the resulting reduced variables. Meaning all of these variables are important in explaining results.

A PCA analysis on the 5 result variables of figure 5 shows that 91% of the variance in these variables can be explained through a reduction to 3 variables. 70% of this explained variance is already explained by the first reduced variable, with the other variables both contributing around 15% of the explanation power. Also, all variables are contributing to values in a meaningful way. The weight of each of the variables in the three output dimensions is shown in Table 15.

	Response	Data_sent	Data_correct	Ans_q	Ans_spec
Dim. 1	0.247262	0.535327	0.511183	0.502212	0.372507
Dim. 2	-0.198751	-0.210168	-0.537863	0.338313	0.715943
Dim. 3	-0.910465	0.292720	0.083338	0.198177	-0.197861

Table 15 Influence of response variables on dimensions following the PCA

When examining these specific variables, it can be seen that the first reduced dimension variable is mostly influenced by three criteria: Is data sent, is what is (not) sent correct, and are answers given to the questions. Specific answers and responses also have an effect on this dimension, albeit noticeably smaller. This dimension is interpreted as a somewhat noisy dimension that measures if an organization's response is lawful, since organizations are obliged to send the correct personal data and answer the other relevant questions under both the GDPR and wbp regulation.

The value of the second dimension in Table 15 is mostly dependent on the criteria of a specific answer, with some influence of the answered questions and a big negative influence of correct data that is received. This dimension is thus specifically classifying organizations which responses do not fall within the criteria of the first dimension but do go out of their way to respond to answers in a specific nature. This dimension is therefore interpreted as measuring an organization's specificity. The third dimension also measures a certain degree of specificity since the huge influence of the response criteria means that it differentiates organizations that did not respond to requests at all from those that did at least send a reply.

Through an interpreted PCA, the five dimensions that measure organizations responses have thus been reduced to two, which can both be described by binomial variables. Since compliance is seen as more important in answering the research questions of this thesis project in a societal perspective, the two variables can be combined into one ordinal value.[2] This transformation is shown in Table 9. The distribution of the response score in percentages of organizations is seen in Figure 12.

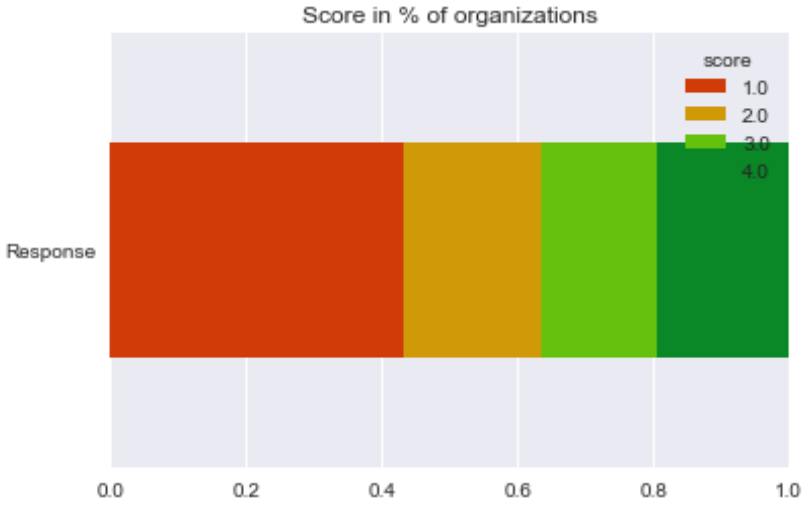


Figure 12. *Response scores of organizations based on the ordinal combination of specificity and compliance.*

		specificity	
		High	Low
Compliance	Yes	4, Excellent response	3, Adequate response
	No	2, Partial response	1, Failed response

Table 16 Combination of the two binomial result variables into one ordinal response quality variable.

[2] The division of all groups of organizations in these four response categories is described in annex A.5.1. The most important groups that was not correctly captured in the dimensions of table 15 are organizations that do not share personal data, but are still coded as being correct in not sending data according to the data subject and those that did respond, but without any answer regarding the request.

As the figure 12 shows, the “failed” group is the biggest in the dataset, with 41 percent of organizations being labeled both noncompliant and not polite in their handling of subject access requests. On aggregate, 35% of organizations is compliant, with 19% responding excellent. By reducing dimensions, a single variable is constructed that is both easy to analyze and easy to interpret. This is especially helpful for analyzing the effects of the more granular organizational variables.

A.5.3 Volunteer analysis

The following graph provides insights in the age and occupation of the 35 volunteers that were recruited for this research. With an average age of 27 years and 82% of volunteers with the occupation “student”, the assumption made in chapter 3 that most of the volunteers would be young students was correct. The relationship between these variables is further demonstrated in figure 13. The volunteer group also includes an overrepresentation of males, with 73 percent of the sample. Figure 13 demonstrates that this has no clear relationship with the age of volunteers.

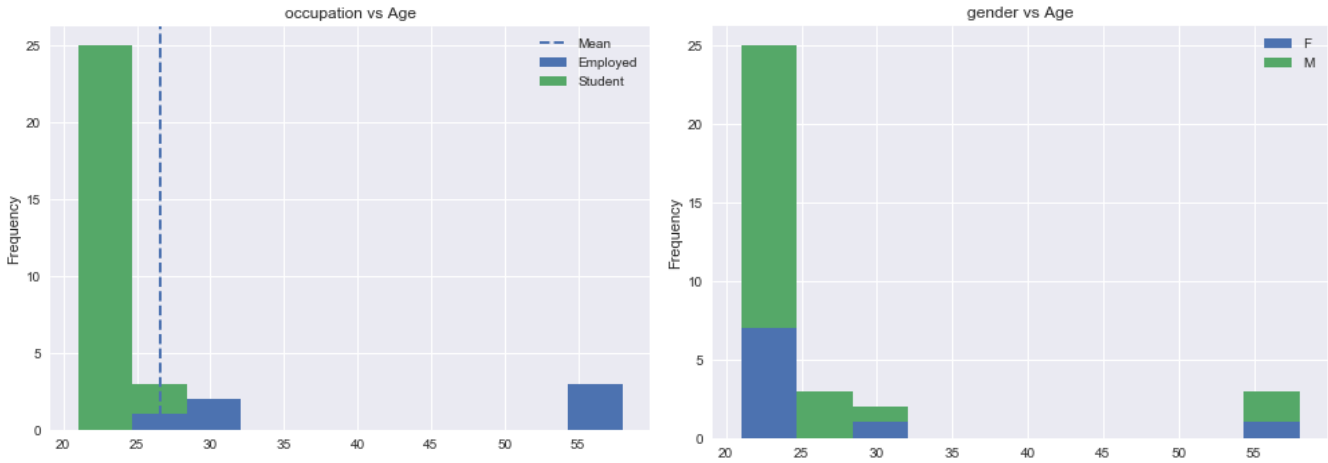


Figure 13 *Histogram of volunteers’ age, color coded by occupation (l) and gender (r)*

A qualitative assessment of the 2000 organizations that the volunteers collectively identified as probably in possession of their personal data provided insight into sectoral differences. Older volunteers were more likely to identify organizations in financial, utility and public sectors, while organizations identified by younger volunteers were more likely to be technology oriented. This means that older volunteers on average provide more interesting organizations for data collection, as defined by the research scope detailed in section 3.2.3. The difference was not large enough to prevent the inclusion of enough relevant organizations in the data collection process.

Further relevant differences between the nature of identified organizations were not identified in the dataset. Differences between identified organizations do seem to exist between genders, but these were mostly limited to specific retail services, and thus did not lead to a challenge in sampling all categories of organizations in the research. On average, volunteers suggested to include just over 60 organizations, which proved to be more than enough for the researcher to pick 10 interesting choices for each volunteer. The distribution of this organization number is presented in figure 14. Its shape is probably influenced by the consent form, which included lines to write down 57 organizations on 3 separate pages.

Volunteers choose to label an average of 14% of these organizations as confidential, with a large group of volunteers labelling none and two volunteers labeling over 40% as such. These confidential organizations were not used for research. Organizations labeled as confidential were almost exclusively in possession of

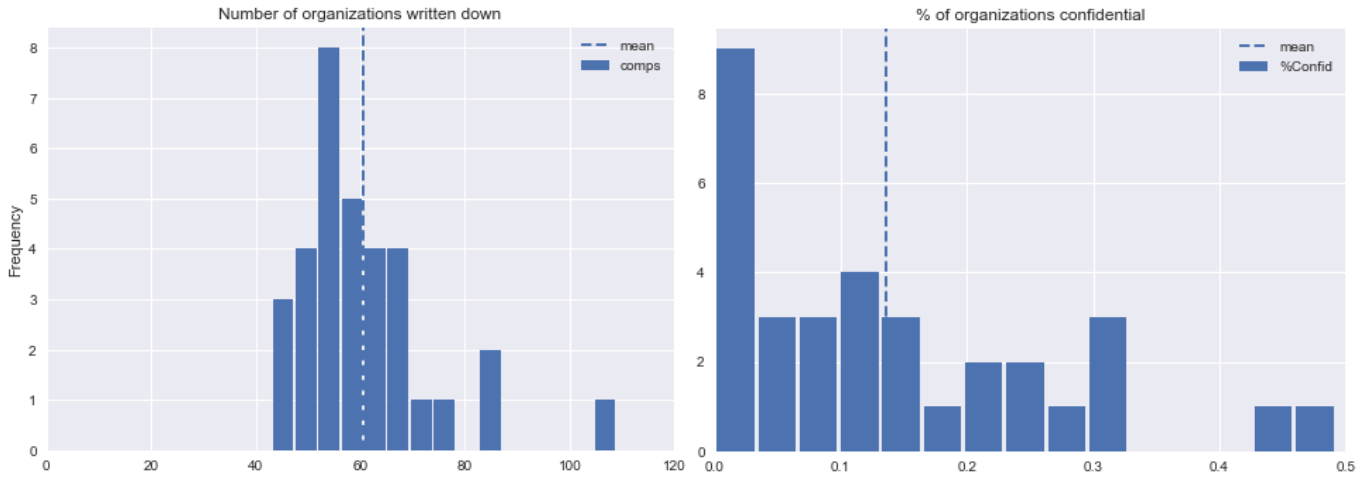


Figure 114. *Histogram of the number of organizations and percentage of these labeled confidential.*

healthcare or financial personal data. The large number of volunteers that did not label any, or just a small rate of organizations as too confidential for research meant that enough organizations that did process this data could still be included in the research.

Volunteers also often asked to exclude their employee or ex-employee from the data collection and "not bother them" with the access requests. This does not mean that volunteers were not interested in the data that was collected on them. Rather, volunteers often considered the formal nature of access requests as inappropriate to communicate with data controllers. Some volunteers that initially gave permission to contact (former) employees on their behalf retracted this later, with one volunteer stating that this decision was made after understanding how much of a burden the request would put on his/her colleagues.

Figure 15 and 16 show the relationship between both the number of organizations and the confidentiality % and a subjectively constructed "closeness factor", describing the relationship between the volunteers and the researcher on a nominal scale where 1 means a very close relation and 4 means a relative loose personal relationship between the researcher and the volunteer. Although the sample size of the data and subjectivity of the factor make any statistical analysis meaningless, the figures seem to suggest a relationship between "closeness" and both other factors. This relationship would make sense, since people intuitively trust those closer to them with more personal data.

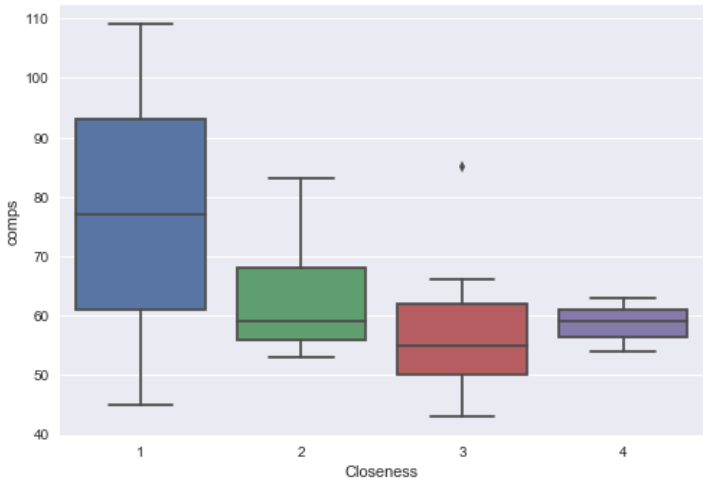


Figure 15 Boxplot of the relationship between volunteer's closeness factor and number of organizations submitted for research.

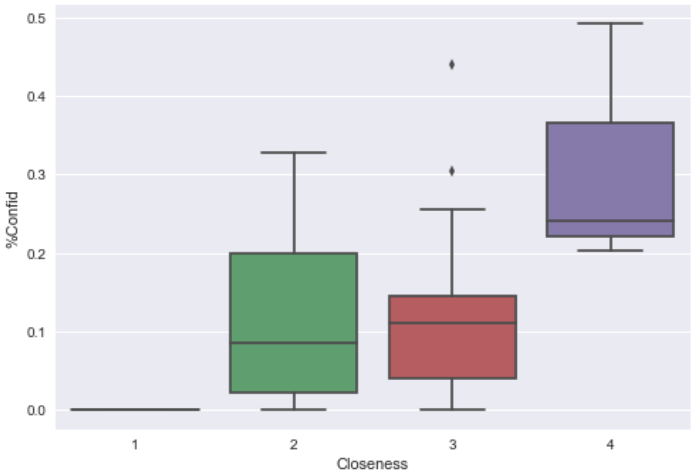


Figure 16 Boxplot of the relationship between volunteer's closeness factor and percentage of organizations submitted for research labeled as confidential.

A final statistic describing the volunteers that participated in the data collection phase follows from their self-described knowledge and involvement in privacy and ICT. Figure 8 shows a boxplot of volunteer's grades. Aside from a relatively high ICT knowledge, the distribution of knowledge and involvement seems to indicate an equally distributed sample. This should not be seen as a conclusive indication of their real knowledge however, as people's perception of their own skill is often different from their actual knowledge (Tracey, Arroll, Richmond, & Barham, 1997). None of these perceived values was found to influence either the Number of identified organizations or the percentage of organizations that was marked as sensitive by volunteers.

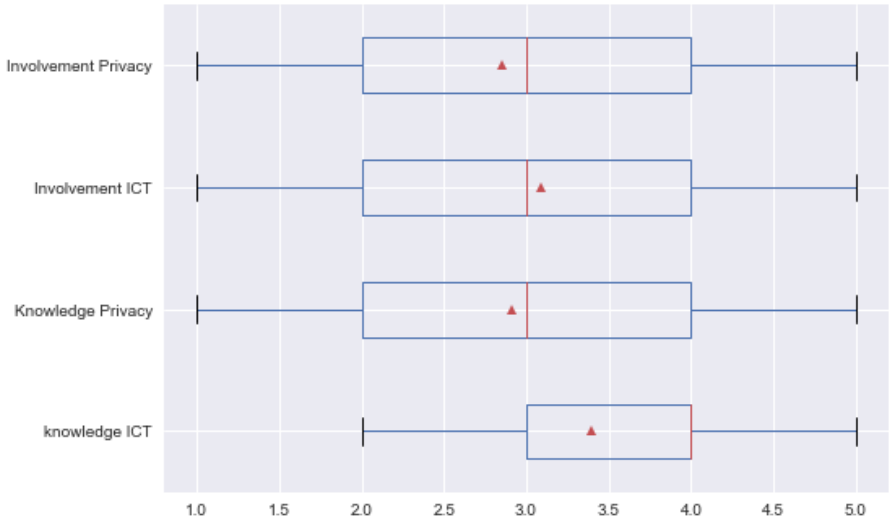


Figure 17. *Boxplots of the ordinal self-perceived knowledge and involvement of volunteers on the topic of privacy and ICT. 1 presents the lowest knowledge/involvement rating and 5 presents the highest.*

Some volunteers retracted their consent for analyzing responses from certain organizations or left the research altogether. When possible, measurements of organization's responses were replaced by those from other volunteers.