Complying with the GDPR in the Context of Continuous Integration

by

Ze Shi Li
B.Sc., University of Victoria, 2018

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

Complying with the GDPR in the Context of Continuous Integration

by

Ze Shi Li
B.Sc., University of Victoria, 2018

Supervisory Committee

Dr. Daniela Damian, Co-Supervisor
(Department of Computer Science, UVic)

Dr. Neil Ernst, Co-Supervisor
(Department of Computer Science, UVic)

Supervisory Committee

Dr. Daniela Damian, Co-Supervisor
(Department of Computer Science, UVic)

Dr. Neil Ernst, Co-Supervisor
(Department of Computer Science, UVic)

ABSTRACT

The full enforcement of the General Data Protection Regulation (GDPR) that began on May 25, 2018 forced any organization that collects and/or processes personal data from European Union citizens to comply with a series of stringent and comprehensive privacy regulations. Many software organizations struggled to comply with the entirety of the GDPR's regulations both leading up and even after the GDPR deadline. Previous studies on the subject of the GDPR have primarily focused on finding implications for users and organizations using surveys or interviews. However, there is a dearth of in-depth studies that investigate compliance practices and compliance challenges in software organizations. In particular, small and medium enterprises are often neglected in these previous studies, despite small and medium enterprises representing the majority of organizations in the EU. Furthermore, organizations that practice continuous integration have largely been ignored in studies on GDPR compliance. Using design science methodology, we conducted an in-depth study over the span of 20 months regarding GDPR compliance practices and challenges in collaboration with a small, startup organization. Our first step helped identify our collaborator's business problems. Subsequently, we iteratively developed two artifacts to address those business problems: a set of privacy requirements operationalized from GDPR principles, and an automated GDPR tool that tests these GDPR-derived privacy requirements. This design science approach resulted in five implications for research and for practice about ongoing challenges to compliance. For instance, our research reveals that GDPR regulations can be partially operationalized and tested through automated means, which is advantageous for achieving long term

compliance. In contrast, more research is needed to create more efficient and effective means to disseminate and manage GDPR knowledge among software developers.

Contents

Supervisory Committee **ii**

Abstract **iii**

Table of Contents **v**

List of Tables **viii**

List of Figures **ix**

Acknowledgements **x**

Dedication **xi**

1 Introduction **1**

1.1 Motivation 4
1.2 Methodology 4
1.3 Research Contributions 5
1.4 Research Publications 6
1.5 Thesis Outline 6

2 Background and Related Work **8**

2.1 Background 8
2.1.1 GDPR: A Privacy Regulation 8
2.1.2 NFR Definition 11
2.1.3 Continuous Software Engineering 12
2.1.4 NFRs and Continuous Software Engineering in Practice 13
2.1.5 Continuous Compliance 14
2.2 Privacy Tools and Methodologies 15
2.3 Current GDPR Challenges and State of Research 16

3 Methodology **19**
3.1 Design Science Methodology 19
3.1.1 Research Setting 21
3.1.2 Problem Characterization 22
3.1.3 Development and Evaluation of Artifacts 22

4 Problem Characterization **24**
4.1 Reliance on Manual GDPR Tests 26
4.2 Limited Awareness and Knowledge of Privacy Requirements 28
4.3 Balancing GDPR Compliance in a Competitive Data Business 29

5 Design Science Artifacts **32**
5.1 Operationalizing GDPR Principles into Privacy Requirements 35
5.1.1 Iterative Development and Evaluation of Requirements as Operationalized Requirements of GDPR Principles 36
5.2 Automated Testing of GDPR Requirements using a GDPR Tool . . . 37
5.2.1 Iterative Development and Evaluataion of GDPR Tool 38

6 Discussion and Implications **40**
6.1 Limited time and motivation inhibit use of continuous GDPR compliance 40
6.1.1 Implications 41
6.2 Insufficient knowledge management impedes privacy awareness and compliance 42
6.2.1 Implications 43
6.3 Managers and developers have sharply different priorities for GDPR compliance 44
6.3.1 Implications 46
6.4 Overconfidence in GDPR readiness reduces the visibility of the state of compliance 46
6.4.1 Implications 47
6.5 Offloading privacy concerns relinquishes compliance control to others 48
6.5.1 Implications 49

7 Threats to Validity **51**

8 Conclusion and Future Work **53**

A Interview Questions Template **54**

B GDPR Tool Scan Results **56**

C Publications **61**

Bibliography **68**

List of Tables

Table 3.1 Participant Role and Experience 22

Table 4.1 Relationship between observed challenges to context at DataCorp. One or more contextual factors (rows) contribute to each specific GDPR challenge (column). These contextual factors and challenges are described in more detail in Chapter 4 25

Table 5.1 Mapping of GDPR Principles to Privacy Requirements 34

Table A.1 Interview Questions Template 55

Table B.1 Number of Infrastructure Resources Scanned by GDPR Tool (W represents Week, For confidentiality purposes, totals are rounded to the nearest 25 and anything below 25 is rounded to 25) . . . 57

Table B.2 Number of Potential GDPR Exposures Identified by GDPR Tool per Infrastructure Resource (W represents Week of Scan, Average represents ratio of exposures per unit of resource) 58

Table B.3 Number of Potential GDPR Exposures by GDPR Tool per Infrastructure region (W represents Week of scan, For confidentiality purposes, totals are rounded to the nearest 15 and anything below 15 is set to 15)) 59

Table B.4 Number of Potential GDPR Exposures Identified by GDPR Tool per GDPR Principle (W represents Week, For confidentiality purposes, totals are rounded to the nearest 50 and anything below 50 is set to 50) 60

Table B.5 Number of Potential GDPR Exposures identified by GDPR Tool per GDPR Recital (W represents Week of scan, For confidentiality purposes, totals are rounded to the nearest 50) 60

List of Figures

Figure 3.1 Design Science Methodology 19
Figure 3.2 Road Map of Research 20

ACKNOWLEDGEMENTS

I would like to thank:

Daniela Damian and Neil Ernst, for their exceptional guidance, support, and teaching throughout this journey. Their incredible guidance has been pivotal to my growth as a researcher.

Trevor Rae, David Johnson, and Dave Cheng, for challenging me and providing thoughtful suggestions propelling me to succeed in my research.

Colin Werner, for being an extraordinary mentor and friend, who exemplified not only the necessary qualities to succeed as a researcher, but also what it means to always carry oneself with professionalism and character.

My parents, for always supporting me through whatever ups and downs.

My grandmother, for amazing meals and always providing me with positive motivation.

MITACs and my collaborating organization, for partially funding this research.

Evil is whatever distracts.
Franz Kafka

But man is not made for defeat. A man can be destroyed but not defeated.
Ernest Hemingway

You miss 100 percent of the shots you never take.
Wayne Gretzky

DEDICATION

To my family for inspiring me to persevere through all the tough times.

Chapter 1

Introduction

Modern internet services often provide people with a trade-off between readily accessible goods and services and the expense of losing full control over one's personal data. To facilitate the convenience of receiving real-time location based services and further improve user experience, users may sacrifice their personal data such as location data [83]. As a result, more user data is being increasingly collected and processed. For instance, whenever a user makes a purchase on an online marketplace, the user's shopping data can be used to help recommend relevant items to the user for future purchases. Similarly, if the cell phone signal is poor in an area of a city, collecting and processing users' connection speeds is a strategy to help telecommunication companies to identify areas for improvement. When user data is appropriately collected and analyzed, both companies who provide goods and services and users can enjoy the benefits of such collaboration. However, notable examples [85, 19] in recent years about the malice and abuse of user data by individuals and organizations have damaged the trust between users and organizations. After the fallout of the Cambridge Analytica scandal, Facebook users particularly in the United States (US), conducted a mass exodus from Facebook [64, 71]. The Cambridge Analytica scandal exemplified the perils of an entity abusing user data for purposes never mutually agreed upon.

Notwithstanding the intentional attacks on users' personal data by a data collector or processor, numerous cases of data hacking or accidental release of personal data have also transpired [31, 63]. In the early 2000s, a large number of people from the United States (US) raised concerns regarding data collection by organizations; over 50% of respondents believed that their right to privacy is being challenged [73]. In Europe, close to 50% of people in Germany did feel that their data was adequately protected [20]. As personal data can flow across geographic boundaries, data protec-

tion is not an isolated initiative. For instance, Ann Cavoukian, the former Privacy Commissioner of Ontario, said in 2013 that, *"Privacy knows no borders: we have to protect privacy globally or we protect it nowhere!"* In short, *privacy* has come to the forefront of the news and government legislation.

When dealing with privacy, public perception may be an important aspect for an organization to consider. As stated, users in the US perceived Facebook in a negative light in wake of the Cambridge Analytica scandal [64, 71]. Constant news reporting of such encroachment of user privacy may hurt a user's desire to continue using a software. However, a large organization has significant resources that may allow the organization to withstand the decrease in trust, whereas a small organization may not have such luxury.

The European Union (EU) took a proactive approach to regulating how organizations deal with user privacy. On May 24th, 2016, the EU officially enacted the General Data Protection Regulation (GDPR) [1]. As a modern pioneering privacy regulation, the GDPR quickly became known as one of the most comprehensive and complex privacy laws in existence [5]. Unlike the 1995 EU Data Protection Directive that the GDPR replaced, the GDPR is enforced in all EU member countries. In addition to being applied in all 28 EU member states, the GDPR states that any organization that collects and/or processes personal data from data subjects in the EU (i.e. any identifiable person) must comply with the GDPR or face dire financial consequences. Any organization reprimanded for GDPR violations can expect to be fined up to 10 million euros or 2% annual revenue for minor violations and up to 20 million euros or 4% annual revenue for egregious violations. Hence, any organization that does not intent on losing significant cash flows to penalties nor the public backlash of violating data subject privacy, should take the initiative to adopt and comply with the GDPR. Given the large scope of the GDPR, the EU gave organizations a two year grace period to prepare for the final deadline. Yet, when the deadline approached on May 25th, 2018, various organizations were drastically unready for the GDPR. In fact, numerous organizations made the decision to shut-off entire operations or trimmed down versions of their systems [6, 2]. For example, on May 25th, 2019, visitors of the popular site National Public Radio (NPR) [6] were redirected to a plain text version of the site if they failed to agree to NPR's new terms of agreement. Other sites such as the Chicago Tribune and Los Angeles Times simply refused any visitor traffic with an EU origin [2]. As the GDPR requires an organization to be compliant *at all times* and *even before* collecting and processing data, an organization's easiest course of

action is feature deletion. In other words, remove features that are not compliant and save the trouble of ensuring the compliance of a process or feature.

However, an organization may accept the inadequacies of its systems and continue to operate as usual, albeit the organization may intend to eventually fix non-compliance areas. Based on initial reports leading up to the GDPR deadline, small organizations seemed to have more compliance challenges than large organizations [80]. Intuitively, small organizations tend to have less resources than large organizations. In consequence, small organizations will likely have less available resources to divert to regulatory compliance than large organizations. In a previous study, startup organizations, which were small and had between 11 to 60 employees, experienced rapid change in their pursuit of finding a consistent revenue stream [47]. One downside to the rapid change is the paucity treatment of non-functional requirements (NFRs). Yet, NFRs, also known as quality attributes, architecturally significant requirements, or "an attribute of or a constraint on a system" [46], are important pillars of a system. Considering the immense size and complexity of the GDPR, the GDPR can be interpreted as a crucial privacy NFR for relevant organizations. If a small organization did not reasonably treat privacy NFRs prior to the GDPR, full compliance may indeed be difficult as neglect or late treatment of NFRs may result in serious financial penalties [70]. Furthermore, public perception of compliance is also important for an organization. Aranda et al.'s [10] work showed that a small organization does not have the luxury of making requirements errors; one error can result in the bankruptcy of the organization.

Another complication to GDPR compliance is the increased use of continuous software engineering in organizations [40], particularly in small organizations. Many organizations are adopting continuous activities such as continuous integration (CI) and continuous delivery (CDE) [40] due to the prescribed benefit of rapidly releasing software to customers and receiving quick feedback from customers [28]. Based on the principles of Agile, Extreme Programming, and Lean, these continuous activities strive to bridge the gap between all facets in an organization such as, development and business, development and operations, development and customers. More importantly, continuous activities strive to continuously deliver *value* to customers [40]. However, one of the more noticeable attributes of a continuous activity like CI, is less emphasis on traditional requirements engineering work and documentation [62]. Empirical studies on Agile projects have shown that testing NFRs such as performance, have been troubling [17]. As the GDPR is a complex and massive set of regulations,

acquiring GDPR knowledge may be difficult, not to mention the additional difficulty of testing through manual or automated means. Unlike large organizations, who can enter new markets or weather a calamitous hit to reputation, a small organization's survival may depend on achieving compliance. Hence, the GDPR and its corresponding privacy NFRs are paramount attributes for any organization that operates on personal data from EU citizens.

1.1 Motivation

In addition to the scores of organizations not compliant by the GDPR deadline [42, 77], numerous organizations are still not GDPR compliant [61, 27], despite the GDPR existing in full force since 2018. The current state of literature often relies on surveys or interviews to study compliance challenges [80, 69]. Existing literature on the GDPR often exists in the analysis of GDPR regulations and potential effects on users or organizations [84, 48]. Small organization as small and medium-sized enterprises (SMEs) represent 99% of all businesses in the EU [4] and seem to experience more challenges as indicated by early reports [80]. As a result, there is a great need explore SMEs to have a detailed understanding of the compliance practices and challenges experienced in an organization. Additionally, no studies have comprehensively explored practices and challenges of GDPR compliance for a small organization practicing continuous activities, which includes CI. Our research findings, which are beneficial for both practitioners and researchers, fulfills this gap in research with an in-depth exploration of an organization's compliance practice and challenges.

1.2 Methodology

Using design science research (i.e. based on Hevner et al. [52]), we conducted an in-depth mixed methods study using an ethnographic informed approach, including participant observation and interviews. As part of design science research, both the business problems that we identified and design science artifacts developed to address these problems pertained to our collaborating organization. Our collaborating organization, DataCorp, is a small organization that primarily operates in the data industry (i.e. the organization receives a large percentage of revenue from data). DataCorp must comply with the GDPR largely due to the DataCorp's collection of large amounts of data from devices around the world, including the EU. Additionally,

DataCorp practices CI, which is part of the context that may effect GDPR compliance practices and challenges at the organization.

Following the guidelines of design science research, our research has three major elements: problem characterization, development of artifacts, and evaluation of artifacts. Since we began the research without having a full context of DataCorp's GDPR compliance, the problem characterization step (explained in more detail in Chapter 4) allowed us to first understand and become acquainted with DataCorp's work and processes. More importantly, the problem characterization emphasized identifying a relevant business problem for DataCorp [52]. Through interviews, observations, and being a "member" of the organization, we gained a strong grasp of the organization challenges and understanding of their problems, which allowed us to contextualize DataCorp's business problems. In other words, we became familiar with the challenges that prevented DataCorp from easily and effectively achieving compliance. Through the development of artifacts step, we iteratively designed and developed two artifacts with the intention to help alleviate some of these challenges. The first artifact is a set of operationalized GDPR requirements; specifically, a set of privacy requirements derived from the GDPR that can be automatically verified. The second artifact is an automated GDPR tool that facilitates the testing of these operationalized GDPR requirements and helps the organization identify GDPR exposures. Finally, to validate that the two design science artifacts resonated with DataCorp's challenges, the artifacts were iteratively evaluated.

1.3 Research Contributions

This study provides five meaningful contributions for both researchers and practitioners. This study:

1. presents a detailed exploration on the practices and challenges of GDPR compliance in DataCorp; specifically, a mapping between context and compliance challenges is provided

2. presents a list of operationalized GDPR privacy requirements that are important to our collaborating organization and derived from three GDPR principles: integrity and confidentiality, data minimization, and storage limitation

3. demonstrates how GDPR privacy requirements can be operationalized in an automated GDPR tool

4. provides empirical data of continuously using an automated GDPR tool to raise awareness about potential GDPR exposures and obstacles of continuous compliance

5. describes five potential hindrances to DataCorp's ongoing GDPR compliance with five implications for researchers and five implications for practitioners

1.4 Research Publications

The aforementioned research methodology and contributions culminated in the following publications:

1. Ze Shi Li, Colin Werner, and Neil Ernst. "Continuous Requirements: An Example Using GDPR," *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, Jeju Island, Korea (South), 2019, pp. 144-149.

2. Ze Shi Li, Colin Werner, Neil Ernst, and Daniela Damian. *GDPR Compliance in the Context of Continuous Integration.* In Review at Transactions of Software Engineering (TSE).

1.5 Thesis Outline

This thesis is organized as follows:

Chapter 1: Introduction provides the motivation for the research, as well as the research methodology and contributions.

Chapter 2: Background and Related Work provides the context of the research and existing work in related areas of this work.

Chapter 3: Methodology details the research process. Specifically, steps taken to define the problem space, as well as developing and evaluating the research artifacts.

Chapter 4: Problem Characterization details the main GDPR compliance challenges that we identified in our collaborating organization.

Chapter 5: Design Science Artifacts describes the iteratively developed and evaluated artifacts that aim to address problems identified in the problem characterization.

Chapter 6: Discussion and Implications describes ongoing challenges to GDPR compliance is specified. In particular, we detail the significance of these ongoing challenges and what it means for practitioners and researchers.

Chapter 7: Threats to Validity describes the limitations as well as the threats to validity of this study.

Chapter 8: Conclusion summarizes this study and provides suggestions for future work.

Appendix: A provides our interview questions template.

Appendix: B provides the scan results of our GDPR tool.

Appendix: C provides the publications that arose from this study.

Chapter 2

Background and Related Work

In this chapter, we first provide an overview and definitions of the topics that form the context of our research, namely, GDPR, NFRs, and continuous activities. Thereafter, we summarize relevant research by inspecting privacy methodologies and tools that pertain to the GDPR or may be used to assist GDPR compliance. Finally, we discuss the current state of research regarding GDPR challenges and practices.

2.1 Background

2.1.1 GDPR: A Privacy Regulation

The GDPR became law on May 24th, 2016 with broad coverage and widespread impact. As the EU represents the world's second largest economic zone and hundreds of millions of people, the sudden change in laws governing the treatment of personal data severely impacted countless organizations who inexplicably had a compliance deadline. The GDPR replaced the two decade old 1995 EU Data Protection Directive with updated privacy regulations [1] that reflected modern technological capabilities. Instead of being a "directive" that prescribed each EU member state to implement its own privacy legislation, the GDPR unites the EU under one umbrella privacy law. Despite the GDPR encompassing more stringent requirements, the GDPR provides an organization with the luxury of adhering to a single law.

The GDPR has six main data processing principles [5]:

1.) lawfulness, fairness, and transparency,

2) purpose limitation,

3.) data minimization,

4.) accuracy

5.) storage minimization, and

6.) integrity and confidentiality.

Accountability is also listed as an additional principle that requires an organization to take appropriate privacy measures and demonstrate compliance [3].

For instance, the storage minimization principle requires an organization to not store a data subject's data for any longer than necessary. An organization complying with the GDPR must ensure that the organization meets all of the principles. As each principle is written at a high level, an organization that wants to realize a principle may need to refine and develop each principle into smaller, more specific requirements. Not only must an organization remain compliant at all times, but also achieve and demonstrate compliance before data collection and process even begins. Hence, once the GDPR grace period ended on May 25th, 2018, a non-compliant organization could no longer *legally* collect nor process personal data. Consequently, an organization does not have the leisure to ask for retroactive permission.

In addition to the GDPR principles, the GDPR also grants a data subject a plethora of rights. These rights include the right to erasure, right of access, right to restrict processing, right to data portability and various other rights. Any specific right may include multiple tangible requirements. For instance, *right to erasure* prescribes a data subject's right to request removal of his or her data at any time and the corresponding organization must oblige within a reasonable time frame [5].

Another important aspect of the GDPR is the shared responsibility placed on a controller (i.e. entity that determines purpose of data collection and collects data) and its processors (i.e. entities that processes data on behalf a controller or told what data to collect) [3]. A processor not only must adhere to the data processing principles of a controller, but also keep a record of its processing activities on behalf of a controller. If an organization's data subjects are its employees, then the organization is also considered a controller. As such, an organization must be wary of its business relationships between other organizations as any other business partner that handles personal data is closely scrutinized by the GDPR. Regarding the right to erasure, a controller must ensure its processors delete all instances of a user's data upon request, even if the data is part of long term archives or backups [67]. Therefore, despite an organization receiving a two year adoption period to prepare for the GDPR, the GDPR deadline may be overwhelming for an organization if the organization did not take time to get prepared.

As stated in Ann Cavoukian's 2013 quote, *"Privacy knows no borders: we have to protect privacy globally or we protect it nowhere!"*, geographic barriers no longer prevent an organization from sharing user data across countries and regions. Accordingly, the GDPR is only the first of many stringent privacy regulations to come. Although federally the US has not moved to update its privacy laws, many states are establishing their independent privacy laws. In particular, privacy regulations that were enacted include the California Consumer Protection Act (CCPA)[1], Vermont's Data Broker Regulations[2], and Improve Electronic Data Security Handling (SHIELD) Act[3]. A law pending legislative passage include the New York Privacy Act (NYPA)[4].

Vermont's Data Broker regulations is the among the newest privacy laws in North America and regulates how data trading organizations collect and sell user data, while also adding requirements for data protection. Vermont's law may have only affected a small number of organizations [60], but its action represents the new trend of tough privacy laws in North America.

Modelled similarly to the GDPR, the CCPA has received widespread attention due to its implication on the most populous US state and one of the world's largest economies. Therefore, the CCPA has large sway over users and organizations. For instance, an organization based in Canada is unlikely affected by the new Vermont privacy law, but there is a more significant likelihood that the organization must consider the CCPA. One noteworthy addition in the CCPA is the requirement that an organization must provide a "Do Not Sell My Personal Information" link to users on a software's home page [56]. Ultimately, CCPA strives to guarantee a user has full control of his or her personal data. The CCPA also has stronger penalties than the GDPR as an organization can be fined thousands of dollars per violation per user [75]. For example, if an organization was found to violate the privacy of millions of users in California, under the CCPA, the organization could expect a potential fine of over 100 billion dollars [75]. One opportunity for a small organization to avoid the CCPA is that revenues are less than 25 million dollars per year, but only if the organization's majority revenue stream is not from selling user data.

In contrast, New York's SHIELD Act is mostly a breach notification and safeguard

[1] `https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375`

[2] `https://legislature.vermont.gov/Documents/2018/Docs/ACTS/ACT171/ACT171\%20As\%20Enacted.pdf`

[3] `https://www.nysenate.gov/legislation/bills/2019/s5575`

[4] `https://www.natlawreview.com/article/new-york-considers-aggressive-consumer-privacy-law`

expectation regulation[5], but nonetheless, both large and small organizations must comply. Moreover, the SHIELD Act conspicuously serves as the precursor to the more comprehensive NYPA. The NYPA[6] is arguably even more restrictive than the CCPA and GDPR. Under the NYPA, individuals have the right to sue organizations for privacy violations and organization of all sizes are affected [57]. For users, the right to sue is a direct improvement from the GDPR, where a user could only report privacy infringements to a data protection agency. Yet, the first iteration of the NYPA did not pass the New York State Legislature, but as data privacy becomes more prominent, similar regulations will likely be reintroduced in the near future. The four privacy regulations discussed in this section exemplifies the increased privacy scrutiny in the US. Hence, it is in the best interest of an organization to comply with the GDPR as comparable and potentially even more daunting regulations may soon be common around the world.

2.1.2 NFR Definition

Non-functional requirements (NFRs), also known as quality attributes or architecturally significant requirements, can profoundly a system's architecture [16]. Various studies have tried to characterize NFRs [16] [45] [46], but NFRs in practice can still be "difficult to enforce during development, and very hard to validate" [21]. Unlike a functional requirement (FR) that focuses on "what" behaviour a system will perform, an NFR characterizes "how" a system will perform that behaviour. Users or customers often have little concern for NFRs [24] as functionality, is typically more apparent than a software's qualities [18] [29]. In contrast, architects are more concerned with NFRs than customers or users [9].

A user may consider specific NFRs such as performance or privacy when deciding whether to use a software or service. For instance, a point of sales system that processes user transactions is unlikely going to attract many users if the system is immensely slow. More importantly, if the system takes no precaution to protect user privacy and publicly discloses transaction data, the system would likely face an abundance of lawsuits not to mention severe loss of public trust. Similarly, developers and managers working in the trenches of software development may not specifically define a suite of NFRs or even a single NFR, but look for distinct qualities in their

[5] https://www.nysenate.gov/legislation/bills/2019/s5575

[6] https://www.natlawreview.com/article/new-york-considers-aggressive-consumer-privacy-law

software. As a privacy requirement, the GDPR and its regulations have significant effect on an organization's software and guide *how* the software is designed, implemented, and operated regarding the treatment of personal data. As an organization must contemplate the implication of each regulation on the design and development of the organization's software, each GDPR regulation can be interpreted as one or more privacy NFRs.

However, despite the number of abundant testing frameworks or tools for NFRs, testing NFRs is still difficult [21]. Additionally, one study shows that manual methods is the default strategy employed to test NFRs [25]. Manual testing may be preferred due to its low upfront costs and adaptability to different situations. However, the downside for over relying on manual testing could be its potential for lack of consistency and significant demands for time [82]. In response to the difficulties of testing an NFR, it is plausible that an organization decides to proceed completely or temporarily without tests. Neither options is ideal as late testing NFRs was found to be less successful than early testing [68].

2.1.3 Continuous Software Engineering

For the purposes of this paper, continuous activities refers to the term continuous software engineering coined by Fitzgerald and Stol [40]. Continuous software engineering includes a wide breadth of activities ranging from continuous planning, continuous experimentation, CI, CDE, continuous deployment (CD), and continuous compliance. Of the continuous activities, the ones with most recognition are CI, CDE, and CD. With the advent of agile development, an organization practicing agile has been encouraged to shorten the organization's feedback loop and the interval between releases [13]. Popularized by Fowler [43], CI provides support for an organization to develop more reliable software with potentially fewer bugs, and facilitates the organization's ability to release more quickly. Since Fowler published his description of CI, more organizations are adopting CI as they become aware of the benefits afforded by CI. Fowler also introduced the concept of a "deployment pipeline" [43], otherwise known as a build pipeline, which is the process from a commit to the eventual release of the commit to customers. A deployment pipeline includes elements such as testing, building, deploying, and reviewing code. Through practices like automated unit, and integration testing, automated builds, rapid builds, and automated deployment [43], CI empowers an organization to catch bugs as early as possible to safely release soft-

ware to customers. Hence, the organization may rapidly acquire accurate feedback from customers.

Two additional continuous activities, continuous delivery (CDE) and continuous deployment (CD), build on top of CI's principles [54]. Both CDE and CD extend the principles of CI with further automation. Specifically, CDE adds automated acceptance testing to an organization's deployment pipeline in addition to automated unit and integration testing of CI. Furthermore, a key difference between CDE and CI is that CDE allows an organization to release at any time, albeit a manual trigger is necessary. Release is easy because CDE requires an organization to keep its software in a releasable state at all times. CD takes CDE a step further by helping an organization release software to customers after each commit. Essentially, if a commit successfully passes each stage in an organization's deployment pipeline, a new version of the organization's software will be released to customers. Practicing CI, an organization can realistically release updated versions of its software hundreds of times a day, but the organization can release with confidence because CI is supposed to help minimize the risk of releases and decrease the difficulty of rolling back software. Hence, CI can facilitate an organization's ability to rapidly release a product that is continuously improved to customers based on quick customer feedback.

2.1.4 NFRs and Continuous Software Engineering in Practice

For the most part, the study of NFRs in the context of continuous activities is an unexplored area. Studies have either focused solely on NFRs or continuous activities. Nevertheless, one study that looked at continuous testing found that NFRs requiring more time to test may be a cause for organization not allocating enough time to test NFRs [65]. This may be due to an NFR being intrinsically contradictory and difficult to validate [30] Moreover, in cases where testing is accepted, the quantity of automated tests is tenuous [58]. Notwithstanding CI's push towards automated testing, an organization practicing CI may still depend on manual testing to test NFRs.

When investigating a group of startup organizations, it was found that NFRs are frequently neglected [47]. Early on in an startup, NFRs "have low priority compared to validating the product idea and the market" [47]. Moreover, important NFRs like, security and usability are not prioritized at first, but become increasingly im-

portant later [47]. As a startup ignores NFRs and take shortcuts to quickly reach the market, technical debt may accumulate. For an organization practicing continuous deployment, it was found that delayed treatment and testing of NFRs may have serious ramifications, as rapid releases of software may allow"resource and performance creep" [74]. Since an organization is likely incrementally improving its software with small updates, insufficient testing of efficiency or performance NFRs will likely lead to a slow degradation of performance over time. In consequence, an organization disregarding NFRs may eventually reach acquire significant technical debt [14], which may lead to rework [38]. At that point, an organization may no longer be able to develop new features or suffer from significant deficiencies in software quality. Privacy is one NFR that may be difficult to prioritize later in a software's life [8]; without initial emphasis on security, later implementation of privacy may also be troublesome due to the intrinsic bond between security and privacy [55]. CD is often practiced by startups, due to its affinity to facilitating quick release of software. However, a trade-off may occur to neglect testing NFRs as these startups perceive that any problems can be quickly resolved in an upcoming release [47].

2.1.5 Continuous Compliance

In addition to the aforementioned continuous activities, continuous software engineering also includes continuous compliance, which is pertinent to both to privacy regulations and NFRs. Continuous compliance is the notion of continuously striving towards regulatory compliance instead of short burst of work followed by long breaks [40]. Continuous compliance is especially relevant to an organization that practices other forms of continuous activities where the organization develops software in short sprints and adheres to a short feedback loop [40]. When practicing continuous compliance, an organization is expected to test for compliance during each sprint. Any non-compliant element is noted and converted into a work task to be added to the organization's backlog and prioritized so that the non-compliant element is quickly resolved. In short, an organization is expected to perform a *complete* compliance review in every sprint as opposed to yearly or bi-yearly. The benefit of continuous compliance reviews is that a non-compliant element of a system can be identified and resolved without much delay. More importantly, a non-compliant element is not expected to "continuously" reappear.

When one considers the demands of the GDPR, particularly regarding the need

to be compliant at all times, naturally, continuous compliance seems like a judicious strategy to assist an organization's GDPR compliance. Given the scale of the GDPR and frequency of testing desired by continuous compliance, relying solely on manual tests is likely bleak, if not impossible. Hence, an organization may be inclined to heavily commit to automated tests for the viability of continuous compliance.

2.2 Privacy Tools and Methodologies

Over the past few decades, the increased integration of privacy and technology has produced privacy enhancing technologies (PETs) and privacy by design (PbD), which aim to increase privacy in software [26]. In particular, PETs use technology to protect the privacy of individuals or a group of individuals [51]. Some forms of PETs include protecting user identities [66] and anonymizing network data [35]. PbD is a more comprehensive concept that not only calls for the prioritization of privacy from the onset of an organization [26], but also during each stage of a software life cycle including planning, development, and operations. However, the extent of PbD's prioritization may be dampened due to situations where "developers are actively discouraged from making informational privacy a priority" [50]. To strive towards optimal treatment of privacy, organizations are recommended to include positive reinforcements and motivate developers to increase their value of privacy [50].

Other privacy methodologies include Deng et al.'s [34] LINDDUN Methodology, which aims to identify privacy threats in a system through analysis of the system's data flow diagram. Since analysis is based on an overview of the system, LINDDUN primarily provides a high level analysis of privacy threats as opposed to specific implementation details [34]. One privacy solution strategy discussed by Deng et al. [34], "removing or turning off the feature is the only way to reduce the risk to zero", was observed in our study and discussed in later sections. We observed that DataCorp shut down potentially concerning elements of its system before the GDPR deadline to decrease risk and hassle. When the risk of an element is excessively difficult to mitigate, removing the element is the safest approach.

Like PbD, other privacy methodologies exist to enhance privacy in an engineering context, but organizational commitment from the inception of a system is often needed as delayed focus on privacy may be too late [49]. Moreover, it is suggested that organization should not view privacy as a zero-sum game but rather organizations can achieve business value from embracing user privacy [22]. Privacy-by-policy and

privacy-by-architecture are two approaches [81] to protect privacy. Privacy-by-policy is the concept of modifying a system to suit privacy, often using privacy policies and user choice as mechanisms. Privacy-by-policy is less reliable and robust, but is frequently adopted by businesses due to its convenience [81], as well as being a popular choice among developers [50]. Privacy-by-architecture is the notion to fundamentally incorporate privacy into a system [81]; user data are anonymous and efforts to exploit user data are futile [81]. Privacy-by-architecture is more reliable, but it has stringent privacy expectations [81]. Ideally, an organization adopts the safer privacy-by-architecture approach, but the approach may not be easily adaptable to a pre-existing system [81].

2.3 Current GDPR Challenges and State of Research

Initial surveys leading up to the GDPR deadline indicated that the number of organizations that would be compliant on time was inauspicious [42, 77]. Some organizations even claimed that achieving compliance may take four years [76]. Hence, one year post GDPR deadline, many organizations are still not GDPR compliant [61]. In some circumstances, organizations even claim that full GDPR compliance is not feasible [27]. Leading up to the GDPR deadline, Sirur et al. [80] observed that larger organizations did not report as many difficulties as smaller organizations. In particular, smaller organizations that did not previously value appropriate security and privacy measures like privacy by design [26] felt burdened by GDPR compliance [80].

Multiple frameworks have been suggested to assist GDPR compliance. Brodin proposes a framework with steps to guide an organization to compliance [23], but the framework is relatively high level and instructions lacked details as to how an organization may implement each step. Similarly, a 6 step approach was proposed to help an organization elicit solution requirements from the GDPR [12]; the appropriateness of the requirements were validated with privacy experts, but the requirements lacked clear cut measurable elements for validation. In contrast, our requirements are more discernible, which allowed us to operationalize requirements into an automated tool.

Coles et al. [32] described a tool supported approach to performing a data protection impact assessment (DPIA), which is one method to demonstrate compliance. Our research focused on a different aspect of compliance, which is helping to achieve

compliance as opposed to demonstrating compliance was achieved. Another study prescribed a framework with 9 steps to help identify privacy risks within organization including risks in infrastructure and business [44]. Holistically analyzing the GDPR, Tikkinen-Piri et al. [84] found 12 ramifications that an organization must be cognizant of and called for more empirical research into GDPR compliance and challenges, for example, in SMEs. Additionally, there have been calls for studying privacy from an organizational perspective [15]. Our study answers this call for further research into GDPR compliance practices and challenges; we found three challenges to compliance and five hindrances to ongoing GDPR compliance.

Although compliance challenges have not been specifically studied in SMEs practicing CI, some work has investigated challenges in other contexts. After interviewing 6 experts involved with implementing the GDPR, Ataei et al. [11] found three compliance challenges related to user interfaces of location based services. One of the challenges was also awareness, but their focus was on raising awareness early in development. In contrast, the challenge of awareness we later discuss refers to awareness throughout the life cycle of an organization and software, not just in early stage development.

Regarding user rights, Altorbaq et al. [7] conducted 10 interviews to formulate guidance for adherence to GDPR data subject rights. Altorbaq et al. found 12 challenges regarding these rights and provided 14 recommendations that may help address these challenges. The recommendations are grouped by stages of a personal information life cycle model created by the same authors. Moreover, the right to data access is a complicated facet of the GDPR as a study with 5 EU insurance companies found 13 challenges related to data access. The challenges were split into four subsets: procedure, protection, privacy, and proliferation [48]. For service oriented SMEs, a study mapped a set of requirements generated from constraints that applied to the studied organization as a result of the GDPR and modified the SME's architecture to satisfy these requirements [53]. In contrast, our work focuses on the efficacy of "operationalizing" privacy requirements derived from a set of GDPR principles, in an automated tool, and in a CI context.

One aspect of the GDPR that frequently frustrates organizations and researchers is its ambiguity, but Cool [33] explained that the GDPR is intentionally vague because it must anticipate future technologies. If the GDPR was specific, the GDPR would have to change after each new innovation or advance in data process or collection. Moreover, since the GDPR applies to all EU member countries, the GDPR

must satisfy each EU county. GDPR regulations may be ambiguous so that each data protection agency (DPA) can interpret each regulation for the DPA's context. Ringmann et al. [72] defined technical requirements that served to help make a software compliant. While the requirements may be mapped directly from the GDPR, the requirements are quite generic as the authors wanted the requirements to apply to as many organizations as possible [72].

However, understanding the theories regarding compliance only represents one aspect of complying in practice. Static code analysis is suggested as a method to identify potential GDPR exposures, but static analysis is limited to code not other candidates for non-compliance like infrastructure or policies [39]. Moreover, to ensure that an organization always abides by the GDPR, continuous compliance may prescribe the necessity for continuously applying a multitude of automated security and privacy tools. For example, our GDPR tool helps raise awareness about possible GDPR exposures. IBM Security Guardium Analyzer aims to analyze a database and identify personally identifiable information [79]. Likewise, Hewlett Packard Enterprises' GDPR Starter Kit aims to help classify data[7]. Ultimately, flagging potential non-compliant candidates is only one step towards continuous compliance. As we will discuss later, continuous compliance requires a coerced effort from an organization.

[7]https://www.hpe.com/us/en/newsroom/news-advisory/2017/05/hpe-software-launches-gdpr-starter-kit-to-expedite-and-simplify-compliance.html

Chapter 3

Methodology

In the following sections, we describe our research approach, including the methods we used.

3.1 Design Science Methodology

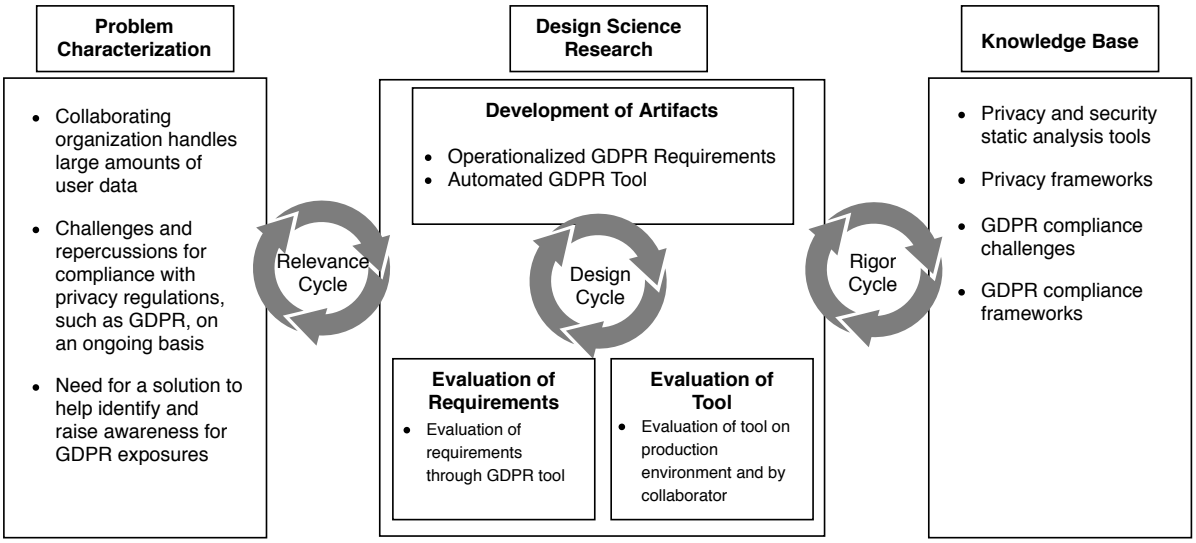

Figure 3.1: Design Science Methodology

The two driving forces of our research were the gap in knowledge of compliance practices and challenges of small organizations practicing continuous activities, as well as our collaborating organization's urgency to achieve GDPR compliance. With operations in the EU, our collaborator inherently expressed interest in researching

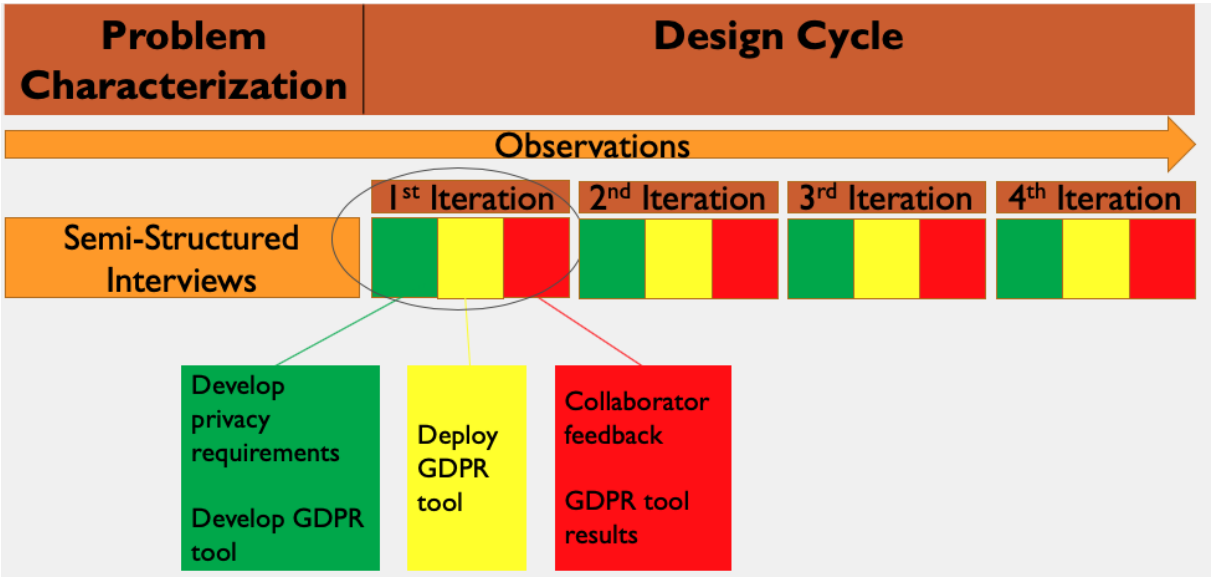

Figure 3.2: Road Map of Research

effective means to become and remain GDPR compliant. Similarly, we looked to broaden the knowledge on applying GDPR to a small organization in a commercial setting in the context of continuous activities. Furthermore, we were motivated to research the organization's challenges and whether practical solutions could mitigate some of those challenges.

Through design science research [78] [52], that involved a mixed-methods approach spanning over 20 months, we conducted ethnographic informed methods including participant observation and interviews. We acquired first hand insight on compliance practices and challenges experienced by an startup organization and studied how an automated tool may help an organization's compliance. We chose design science because it emphasizes the importance of finding relevant problems in the investigated organization and producing artifacts to reduce the burden of these problems. Figure 3.1 depicts the elements of our design science methodology. In particular, the left part of Figure 3.1 depicts the findings of our problem characterization, which serve to uncover important and relevant business problems [52]. Although we began our research with very little knowledge about our collaborator, design science allowed us to become acquainted with the organization through the problem characterization step. The design science artifacts produced in our research must not only be relevant for our collaborator, but also rigorously evaluated [52]. Similarly, Figure 3.2 illustrate the road map of the study.

3.1.1 Research Setting

DataCorp[1] is a startup organization that operates in the data industry. During our study, DataCorp experienced immense growth, whereby its employees increased more than three fold. DataCorp's business involves collecting data from worldwide users. Every day, millions of data points may be shared by millions of users, many of whom are EU users. Since the GDPR prescribes GDPR compliance from any organization that collects personally identifiable data from EU citizens, DataCorp had the obligation to become compliant by the GDPR deadline. As a precautionary measure to protect user privacy, DataCorp pseudo-anonymized data when data is received from users. For development, DataCorp uses CI tools like Jenkins to automate software build and deploy software to production. After code is committed and pushed to source control, DataCorp's deployment pipeline builds the code and runs automated tests against the code, if pertinent tests exist. Furthermore, DataCorp heavily relies on third-party services and tools hosted in the cloud for storing and working on data. Due to the nature of DataCorp's business, DataCorp's partners are split into a few different categories: 1) customers who receive data from DataCorp 2) partners who enable DataCorp to collect data 3) third-party services and tools that provide the basis of DataCorp's infrastructure.

As part of our research, the author of this thesis led a mixed-methods approach including participant observation and interviews, whereby the author became a member of DataCorp and acquired first hand understanding of DataCorp's activities. The author spent one to two days per week in DataCorp's offices. To acquire a reasonable perspective of DataCorp's work, the co-author participated in meetings, such as planning and retrospective meetings and performed tasks such as documenting data flow. We also received access to some of DataCorp's source control repository, project management tools, and infrastructure hosted in the cloud. Furthermore, we interacted with employees, conducted interviews, as well as learning and observing the organization's processes. Figure 3.2 helps to depict the research process. Observations and interviews occurred during the problem characterization, but observing and being part a member of the organization continued throughout the study.

These types of activities facilitated our increase in awareness on how DataCorp planned work, developed code, tested software, and types of tools used to support DataCorp's work. In addition, analyzing project management tasks gave us insight

[1]The real name of the organization was changed for confidentiality purposes

Table 3.1: Participant Role and Experience

Id	Role	Time in Organization
P1	Developer	Less than 5 years
P2	Developer	5 or greater
P3	Manager	5 or greater
P4	Manager	5 or greater
P5	Developer	5 or greater
P6	Developer	Less than 5 years
P7	Developer	Less than 5 years
P8	Developer	Less than 5 years
P9	Manager	5 or greater

on the type and distribution of tasks, as well as the amount of preparation conducted for GDPR compliance. Ultimately, our study strengthened our understanding on the practices utilized by DataCorp for compliance and active challenges that hinder the organization's compliance ability.

3.1.2 Problem Characterization

The problem characterization step of our research sought to understand the challenges experienced by DataCorp. Hence, we participated in meetings, conducted interviews, observed discussions, and conversed with other employees. During the initial eight months at DataCorp, we participated in at least six meetings, conducted interviews with nine employees, observed numerous discussions, and conversed with essentially every DataCorp employee. Based on our problem characterization, we identified relevant problems in the organization and found potential causes of these problems. We noticed three main challenges that DataCorp experienced, which we will elaborate in chapter 4.

3.1.3 Development and Evaluation of Artifacts

As shown by Figure 3.1, to mitigate the difficulties found in our problem characterization, we produced two iteratively developed and evaluated artifacts: operationalized GDPR requirements and an automated GDPR tool. The development of our artifacts was heavily influenced by the compliance challenges we observed at DataCorp especially the challenges of manual testing and awareness. Hence, it was paramount

that DataCorp provided guidance and feedback in the evaluations of our artifacts to ensure that these artifacts are relevant and practical to DataCorp.

Chapter 4

Problem Characterization

Our design science research first establishes the relevance of our research to an actual business setting at DataCorp. As part of problem characterization, we interviewed nine employees, which consisted of developers and managers. Table 3.1 lists each interviewee's primary role and time spent in DataCorp. Due to our ethics guidelines and NDA signed with DataCorp, we anonymized each interviewee. A developer represents someone who mostly works in development, testing, or operations. In contrast, a manager represents someone whose primary focus is managing developers or other employees. A "manager" may still perform development tasks as DataCorp is a startup and employees often have multiple responsibilities. Table A.1 from the appendices section lists the template of questions that we asked each interviewee. Since we conducted observations and interviews, we could corroborate our findings to define the problem instance. During the interviews, we also ran an survey where each interviewee was asked to prioritize NFRs based on a list of thirteen NFRs. The survey entailed two iterations. The first iteration involved ranking each NFR based on an interviewee's role, whereas the second was from the perspective of the business.

We identified three main challenges in DataCorp that hinder GDPR compliance:

1. reliance on manual GDPR tests,
2. limited awareness and knowledge of privacy requirements,
3. and balancing GDPR compliance in a competitive data business

Table 4.1 maps our observed context and circumstances to the challenges. We describe the challenges in more detail in the following subsections.

Table 4.1: Relationship between observed challenges to context at DataCorp. One or more contextual factors (rows) contribute to each specific GDPR challenge (column). These contextual factors and challenges are described in more detail in Chapter 4

.

		Challenges						
		Awareness and Knowledge			**Testing**	**Business and Workflow**		
		Understanding the GDPR	*Becoming aware of new regulations*	*Educating users*	*Manual testing of privacy requirements*	*Checking for GDPR compliance*	*Long term GDPR compliance*	*Ensuring compliance from customers and processors*
Context	Number of GDPR Regulations	X						
	Ambiguity of GDPR	X	X					
	Lack of legal training	X						
	Lack of privacy experience	X						
	Consultants advice fron experts	X						X
	Nature of business			X			X	X
	Size of organization		X		X	X	X	X
	Lack of time				X	X	X	
	Increased growth of infrastructure and data				X	X		X
	Data subject rights granted by the GDPR				X			X
	Making existing systems compliant					X	X	X
	Lack of shared understanding					X	X	

4.1 Reliance on Manual GDPR Tests

DataCorp over relies on manual tests for verifying GDPR compliance. Overcompensating on manual tests leads to significant strains on time, which can invoke time pressure on an employee tasked with verifying compliance. However, continuous growth of DataCorp's infrastructure as a result of DataCorp maturing further intensified the challenge with manual testing given low allocatable time.

When our research began, DataCorp was much smaller in size (i.e. three times fewer total employees) and we observed employees frequently performing a multitude of responsibilities and stressed by time constraints. Hence, we often heard employees say "I would...but I have no time" (P2) or "I wish I had more time" (P6). DataCorp uses some forms of automated tests, like using an automated crawler to verify that DataCorp's privacy policy in an App store is consistent with DataCorp's intended privacy policy. Yet, manual tests are still the predominant strategy to test privacy requirements. As stated by P6, *"Privacy requirements are not automatically tested. Mostly conducted through manual means"*. If a privacy requirement stipulated a stoppage of data collection for a specific data parameter, a developer would need to manually check a database to verify the data parameter was no longer collected by the organization's system.

However, manual tests are laborious, error prone, and time consuming [36]. It is easy for a developer facing time pressure to check the wrong database or run the wrong query. Regardless, running an erroneous test would hinder the organization's compliance effort as any erroneous result produced can snowball into a future privacy requirement rework or retest. To check that a privacy requirement still applies after every change to a database, a developer would have to conduct the same type of manual test after every code change.

As DataCorp matures, its infrastructure and data also experiences immense growth. DataCorp cannot continue its manual approach to testing privacy requirements; either developers are redirected from other work or system elements are "assumed" to be GDPR compliant. In particular, DataCorp verifying the GDPR compliance of DataCorp's infrastructure is particularly time-consuming. Manually finding GDPR exposures has the potential benefit of a human interpreting a subjective scenario, but manual testing is slow. Therefore, we found the over reliance on manual testing likely unsustainable for long term compliance.

DataCorp relies on a multitude of third party services like Amazon Web Services

(AWS), Google Cloud Platform (GCP), and Azure; the organization has thousands of infrastructure resources hosted by many third party services. For instance, DataCorp hosts more than 50 databases and over one hundred servers on a single third party service. It is arduous and tedious for a developer to manually review all those databases. Moreover, the quantity of resources also experienced rapid growth; the number of servers on one service increased 14% over a 5 month period. Hence, a developer tasked with uncovering a GDPR exposure in the DataCorp's entire infrastructure may require substantial time.

Data subject rights granted by the GDPR also reduces the allocatable amount of time at DataCorp. Even if a right is onerous for an organization to comply, the organization must abide. For example, a user can request an organization to provide all existing data about the user. An organization must terminate data collection and delete a user's data upon request even if the user gave prior consent to data collection. Thus, soon after the GDPR deadline, DataCorp began receiving emails and requests from various users asking to stop collecting their data. However, DataCorp has a manual termination process that requires individual responses to each user. As explained by P9 *"When a user sends a request to opt out to us, the emails come to me and I have to tell them how to opt out"*, the organization has to respond to each individual user. If the organization receives a plethora of requests per day, P9 would have to help satisfy each user, which may inhibit other important work given each employee's busy schedule.

When we asked our interviewees the question "How long do they think it will take for a developer to find a specific GDPR exposure from a cloud provider like AWS," the answers varied between a day and two weeks. We found that managers and developers differed in their time estimates. Developers often gave auspicious estimates of a day or few days, whereas managers provided more conservative estimates closer to the upper range of weeks.

In general, we found that the excessive use of manual tests for privacy requirements is time-consuming. The challenge is even more compounded when testing GDPR compliance of DataCorp's large and growing infrastructure. Additionally, aspects of GDPR requirements and collaborator's compliance approach further challenged the organization's time commitments.

4.2 Limited Awareness and Knowledge of Privacy Requirements

It can be difficult for DataCorp to properly identify privacy problems, due to the complexity and magnitude of the GDPR and DataCorp's inexperience dealing with privacy regulations. Additionally, the lack of awareness of new privacy regulations may inhibit long term privacy compliance. DataCorp must also manage privacy awareness of users to collect data.

Ideally, each DataCorp's employee is knowledgeable and reasonably understands the GDPR, but attaining a sufficient understanding is difficult. The GDPR consists of ninety-nine articles and one hundred seventy-three recitals [5], but the entire GDPR is written in legal speak. For lawyers, the GDPR may be straightforward, but DataCorp's employees are not well-versed in legal language nor have specific privacy training. Hence, DataCorp's developers are unsure about the requirements dictated by the GDPR, which may prevent effective treatment of a privacy NFR. In addition, GDPR regulations are often ambiguous [33], which further hinders understanding. For example, *"[Evaluating GDPR compliance of tools] is difficult because I am not an expert in the GDPR."* (P1) and *"Interpreting the rules and regulations [was challenging]. The rules weren't clear on what can be collected and what is considered private"* (P9). The inexperience with privacy regulations also reduced the ability to share GDPR knowledge with each other. Furthermore, some GDPR compliance guides lack details or contain inaccuracies, which could create misinterpretations or misunderstandings. Even external consultants often provided contrasting answers to the same question. As such, despite the presence of external experts, knowledge sharing from external sources was not always a boon to DataCorp's GDPR awareness.

For long term compliance, DataCorp should *"Stay up to date with the regulations. Put efforts in research and implement the changes"* (P7). Yet, none of our 9 interviewees could definitively describe an upcoming privacy regulation, but a manager rightly speculated that the US would likely pass privacy laws: *"No [not aware of any new regulations], but US will probably adopt something similar to the GDPR"* (P9). Not knowing an upcoming privacy regulation does not have a direct negative affect on current GDPR work, but awareness of forthcoming privacy NFRs may prevent duplicate work and simplify future compliance adoption. Even staying up to date with the GDPR may be difficult: *"[A large challenge is knowing] changes to the GDPR. Especially minor changes [and amendments] can be difficult for companies to*

find out" (P4).

Another difficulty of managing awareness is educating users on DataCorp's data collection purposes. As explained by P9 *"a user needs to be educated on why we are collecting data"*. The risk of not providing a user with enough explanation is that the user may decline the organization's terms of service or report the organization to a data protection agency; both actions are within a user's data subject rights. Ideally, users receive a terms of service agreement that is readable and trustworthy. As users play a pivotal role to DataCorp's business, DataCorp must sufficiently communicate and inform users about the organization's collection purpose and get consent from users. Therefore, DataCorp not only has the challenge of ensuring an adequate level of internal GDPR and privacy awareness, but also raising the level of awareness from its users.

Regarding privacy work, DataCorp had an unequal distribution of tasks as managers and a few specific developers seemed to receive the bulk of tasks. Hence, many employees felt insignificant impact from the GDPR. This may be a reason why managers generally value privacy more than developers in our NFR survey. In our NFR survey, managers also felt privacy was significantly more important to DataCorp's business than developers. Finally, we also observed that CI may provide a couple advantages for achieving compliance, namely quick release and feedback: *"Through CI, [redacted] can be generated, modified, and fixed within a couple of hours"* (P5) and *"[allows involvement] with external stakeholders"* (P9). However, these compliance benefits may be contingent on employees possessing a sufficient level of GDPR knowledge. For a developer to implement fast changes and receive rapid feedback, the developer has to recognize the expectations of the GDPR.

4.3 Balancing GDPR Compliance in a Competitive Data Business

Due to DataCorp's business needs, DataCorp is naturally affected by the GDPR's stringent regulations. In response, DataCorp shut off aspects of its data collection in the EU as the GDPR deadline approached. Earning the trust of users and receiving consent is paramount to the success of the organization, but even if a user consents, *"there may be a regulator who says we can't collect this data"* (P9). Despite DataCorp's best efforts to justify its data collection and takes adequate steps to safeguard

its systems, a DPA could decide that DataCorp is not allowed to collect data.

DataCorp should *"Stay up to date with the regulations. Put efforts in research and implement the changes"* (P7), for long term compliance of the GDPR, However, staying up to date with the GDPR is difficult: *"[A large challenge is knowing] changes to the GDPR. Especially minor changes [and amendments] can be difficult for companies to find out"* (P4). If the GDPR is ever amended, it is not satisfactory to solely rely on a news article to determine the right course of action for mending non-compliance elements. Instead, the organization needs to make decisions based on GDPR regulations, which both managers and developers have acknowledged are difficult to understand.

Complicating matters for DataCorp is that it is a small organization with many competitors. For instance, *"Staying competitive in terms of [volume of data] we collect and present, while respecting privacy concerns of anonymization"* (P5). To stay competitive against other companies, DataCorp needs to continue increasing the amount of data collected from users. Therefore, DataCorp needs to balance GDPR requirements and DataCorp's business.

Since DataCorp's system already exists, becoming compliant means "upgrading" the entire system to adhere with the GDPR. Aspects of DataCorp system have existed for years, but NFRs are known to be difficult to implement and test late in software development [59]. At this stage, it can be challenging to modify elements that affect the system architecture. P6 admitted that *"building a GDPR compliant product is easier than making a legacy system GDPR compliant"*.

Due to the GDPR's emphasis on shared responsibility between controllers and processors, DataCorp must also vet its partners: *"[It's challenging] making sure that partners who receive data are compliant"* (P3). Furthermore, based on GDPR data erasure policies, if DataCorp receive a request to delete a user's data, DataCorp must also forward the request to *every* partner who received the user's data and ensure that these partners also comply with the user's request. Essentially, DataCorp's compliance is also tied to the compliance of DataCorp's partners.

In addition, knowledge sharing in DataCorp is not yet robust, which may lead to insufficient transparency of work and processes. Particularly, the insufficient sharing of knowledge materialized into instances of developers making misguided assumptions that elements are secure and compliant. For example, a developer explained that GDPR compliance was not a significant concern for him because his work dealt with data that was already pre-processed. The developer's assumed that prior pro-

cesses contained safeguards and checks that would ensure the data is GDPR compliant. While DataCorp is making significant progress regarding compliance and the developer had no harmful intention, the developer assumed that DataCorp is entirely compliant, which is not a fully accurate assumption.

Chapter 5

Design Science Artifacts

Based on the problem characterization step of our design science methodology, DataCorp dealt with three main challenges: reliance on manual GDPR tests, limited awareness and knowledge of privacy requirements, and balancing GDPR compliance in a competitive data business. In response to these challenges, we determined that reliance on manual GDPR tests is the most important and amenable challenge, especially as employees deal with constraints to time and long term manual testing of the GDPR is unsustainable. Over reliance on manual testing is a bothersome challenge particularly pertinent to DataCorp's growing infrastructure. In contrast, the other two challenges are less likely to be directly solved by software solutions.

Therefore, we embarked on constructing artifacts as part of the middle phase of the design science cycle shown in Figure 3.1. The goal is to construct artifacts to help reduce the problem of over relying on manual testing at DataCorp. Our approach is to automate the manual testing to reduce staff effort and improve the repeatability and traceability aspects of compliance. To do this, we first determined which GDPR principles are most amenable to automated testing, within the specific context of our partner. We analyzed DataCorp's infrastructure and the GDPR principles and found three pertinent GDPR principles to DataCorp's infrastructure. These are shown in Table 5.1.

We operationalized these three GDPR principles into specific privacy NFRs that apply to DataCorp. The privacy NFRs were then automated in a custom-built GDPR tool to raise awareness about potential GDPR exposures and continuously verify whether DataCorp's infrastructure satisfies these NFRs.

Over a period of six months, we iteratively developed and evaluated our design science artifacts. Since development of artifacts was heavily influenced by compliance

challenges at DataCorp, it was paramount that DataCorp provided guidance and feedback in the evaluations of our artifacts. Ultimately two design science artifacts were produced: privacy requirements operationalized from GDPR principles and an automated GDPR tool.

Table 5.1: Mapping of GDPR Principles to Privacy Requirements

GDPR Principle	**Privacy Requirement**
Integrity and Confidentiality	A database must be encrypted for integrity
	Each server must exist with a purpose
	Each server without purpose must be removed
	Each server must have a corresponding cloud firewall
	Each server storage must be encrypted
	Each server storage must exist for a purpose
	Each cloud firewall must use secure protocols inbound and outbound
	Each cloud firewall must limit access to reliable sources
	Each cloud firewall must limit outbound communication to reliable sources
	Each load balancer must use end to end encryption
	Each load balancer must use secure protocols
	Each cloud storage resource must be encrypted
	Each cloud storage resource must restrict access from unapproved sources
	Each cloud storage resource must limit modification and deletion from unapproved sources
	Each access management resource must not grant all permissions
	Each access management resource must not grant permissions to all infrastructure resources
	Each router must limit outbound communication to unapproved sources
Data Minimization	Each database must not collect personal data types outside an organization's data collection purpose
Storage Limitation	Each database tuple must not live indefinitely

5.1 Operationalizing GDPR Principles into Privacy Requirements

The GDPR has six main data processing principles 1) lawfulness, fairness and transparency 2) purpose limitation 3) data minimisation 4) accuracy 5) storage limitation 6) integrity and confidentiality [5]. Accountability is another primary GDPR principle, but accountability's purpose is requiring organizations to adhere to GDPR regulations and demonstrate compliance.

Our first design science artifact is our list of privacy requirements operationalized from GDPR principles, shown by Table 5.1. Based on input from DataCorp and our own observations, the integrity and confidentiality principle was the most important candidate to be operationalized (by operationalized, we mean the automatic process of confirming whether an NFR is satisfied). The purpose of this principle is ensuring that the organization is adequately handling personal data, and safeguarding that data from malicious attacks or accidental misappropriation. Hence, this principle requires adequate safeguards in not just the organization as a whole, but also every infrastructure resource. For example, one example of a specific requirement based on this GDPR principle is that databases and servers must be encrypted. This was explained to us by two different employees: *"I added more encryption to the databases"* (P6) and *"[I worked on] disk and storage encryption"* (P2). Prior to our study, some employees were assigned related tasks, but there was no systematic strategy of verifying each infrastructure element, leading to potential privacy exposures in the system. Moreover, Section 4.1 elaborates on DataCorp's extensive cloud-based infrastructure, that makes manual testing of every infrastructure resource an arduous process. As such, we operationalized this principle for a myriad of infrastructure, as shown by Table 5.1.

Our second operationalized principle, storage limitation, as shown by Table 5.1, represents the idea of keeping data no longer than necessary. An organization must ensure that it has a process to remove a datum after a period of time. For example, a datum is automatically removed after a year. We applied the storage limitation principle to databases, which is heavily used by DataCorp to store data. In DataCorp's situation, the multitude of data is ideally automatically removed after a specified time frame.

Similarly, the data minimization principle instills the notion that personal data should only be collected if necessary and relevant to an organization's data collection

purpose. Depicted by Table 5.1, data minimization is our third operationalized principles. As DataCorp collects a large assortment of data and data types, it is onerous for a developer to manually verify whether the organization is collecting more personal data than originally intended. Similar to storage limitation, data minimization also applies to databases.

In contrast, we chose not to operationalize three principles (i.e. lawfulness, fairness, and transparency, accuracy, and purpose limitation), because these principles are more subjective in nature and/or have less applicability for DataCorp. For instance, the accuracy principle prescribes that personal data must be kept up to date and inaccurate personal data is fixed or erased [5]. Personal data collected by DataCorp is pseudo-anonymous; if a data point was inaccurate for any particular reason, DataCorp has minimal ability to identify the corresponding data subject and the data subject is almost certainly not going to be affected. Only DataCorp's partners may be affected as they desire accurate data. Moreover, we did not focus on automating the verification of rights of data subjects granted by the GDPR partially due to the subjectivity of these rights. For example, right of access allows a data subject "if possible" to access his or her data, but there is not a clear distinction on whether access is possible or not. Ultimately, a data protection agency (DPA) has the final authority deciding whether a right "is possible" or not.

We operationalized the principles shown by Table 5.1, as these principles are more stringent and can be more demanding given the magnitude of our collaborator's infrastructure. After all, as DataCorp grows, its infrastructure will correspondingly expand and more compliance testing is needed to ensure DataCorp's continuous adherence.

5.1.1 Iterative Development and Evaluation of Requirements as Operationalized Requirements of GDPR Principles

After determining three relevant GDPR principles, these principles were operationalized into privacy requirements as shown by Table 5.1. Specifically, the privacy requirements were developed based on input from DataCorp and implications of these three relevant GDPR principles on DataCorp's various infrastructure resources. Hence, these requirements are relevant to DataCorp. Moreover, we operationalized the integrity and confidentiality principle for a DataCorp's infrastructure including servers, load balancers, and databases. Resulting requirements include ensuring that an access management resource does not provide a blanket policy that grants unrestricted

access or action and a database is encrypted. Similarly, we applied the storage limitation and data minimization principles to databases, which are heavily used by DataCorp to store its data. For storage limitation, DataCorp automatically removes any long existing data. Likewise, data minimization was refined into the privacy requirement, "A database must not collect personal data types outside an organization's data collection purpose."

Our list of privacy requirements was iteratively refined based on the combined feedback from DataCorp and the results of operationalizing the requirements in a GDPR tool, which we discuss in more detail in Section 5.2. We evaluated each requirement based on two properties: 1) a requirement is important to DataCorp 2) a requirement is derived from a GDPR principle. For example, DataCorp disagreed and felt the requirement "Each load balancer must only use secure protocols" caused our GDPR tool to flag many load balancers that DataCorp perceived as otherwise secure. In response, we revised the NFR to "Each load balancer must use secure protocols" to account for cases where a load balancer listened for both http and https traffic. The previous example represented a requirement that fulfilled the second property, but was not initially important enough to DataCorp as DataCorp felt the requirement was too stringent.

On the contrary, an operationalized requirement is determined to be effective and valid when the requirement comes from a GDPR principle and DataCorp finds the requirement important. For instance, an employee exclaimed, *"It is peculiar that [redacted]...that should have all been fixed a while ago!"* While we do not have the full context on when and how this type of GDPR exposure was originally treated, but the scenario helps exemplify the difficulty of manually fixing and testing for GDPR compliance. Moreover, we also refined our privacy requirements based on lessons learned from external events. For example, when the Capital One breach occurred [31], which largely originated from misconfigurations of cloud infrastructure, we created requirements that applied to access and modification rights.

5.2 Automated Testing of GDPR Requirements using a GDPR Tool

From the list of privacy requirement from Table 5.1, we developed our second design science artifact: a GDPR tool that verifies these requirements and can be executed

automatically on DataCorp's system. Specifically, our GDPR tool entails a series of Python scripts tailored for Amazon Web Services (AWS). In short, our tool checked privacy requirements in various elements on DataCorp's AWS cloud infrastructure. More importantly, our tool provided a vehicle for us to apply our operationalized requirements in practice and validate whether these requirements are reasonable and legitimate.

Assuming GDPR exposures in DataCorp's infrastructure are found, our GDPR tool produces a list with detailed information about each exposure, such as location, name, id, type of resource, and pertinent GDPR principle, which allows a developer to investigate the exposure in more detail. Furthermore, our GDPR tool ran without requiring a human intervening to trigger an execution as Jenkins, a CI tool, triggers the GDPR tool to run weekly. If DataCorp desired, DataCorp could run the tool every minute.

Our GDPR tool was also effected by influences apart of DataCorp. The privacy requirements created from the Capital One Breach [31] were eventually implemented in our GDPR tool. Due to the frequent use of cloud infrastructure to store large amounts of data from around the world, cloud infrastructure is crucial to our collaborator's business. Managing infrastructure is difficult, especially at the scale of many modern applications. We operationalized the integrity and confidentiality principle for access management and our GDPR tool began flagging access management resources that may grant overarching, unnecessary permissions.

5.2.1 Iterative Development and Evaluataion of GDPR Tool

Our GDPR tool serves to realize our privacy NFRs in practice, which allows us to iteratively improve our privacy requirements and the tool itself. In the iterative development and evaluation of our GDPR tool, we received feedback from DataCorp through meetings and discussions, as well as analyzing the results produced by the tool. The feedback helped evaluate the accuracy of the tool as well as improvements such as tool efficiency. For instance, when the GDPR found eleven load balancers of a specific type, we manually verified that there were eleven load balancers from the third party provider. We also modified our GDPR tool to reflect any changes to our list of privacy requirements. Ultimately, the purpose of our evaluation was to ensure that our list of privacy requirements was verified in an automated tool. The tool in turn automatically checked for potential GDPR exposures and provided meaningful

details that can help an employee investigate an exposure. Unfortunately, DataCorp did not consistently create tasks to address identified potential problems found by our tool during the course of study. As shown by Tables B.1, B.2, B.3, B.4, and B.5 in the appendices section, our tool identified potential GDPR problems, but DataCorp did not consistently remedy these problems. It may be that employees were currently limited by time and felt the potential GDPR exposures identified by the GDPR tool are not "severe" enough to cause a drastic penalty if temporarily not investigated and resolved. However, we were encouraged by DataCorp agreement that our GDPR tool's results should have been added to the organization's backlog, but employees have been limited with other work.

Chapter 6

Discussion and Implications

While our operationalized requirements and GDPR tool have helped mitigate some compliance challenges, there are still challenges with respect to ongoing GDPR compliance in an organization such as DataCorp. In particular, we highlight implications for research and practice.

6.1 Limited time and motivation inhibit use of continuous GDPR compliance

Continuous compliance [40] is characterized as automatically checking regulatory compliance after each sprint. Based on continuous compliance, if any non-compliance issues exist, the organization will add the list of issues to the organization's backlog to reduce the chances of the same non-compliance issue continuously recurring. Next, the privacy tasks would be assigned a high priority and resolved in an subsequent sprint. Since our GDPR tool executed on a continuous basis and was automated, we also had the opportunity to explore GDPR continuous compliance at DataCorp.

Our GDPR tool could serve as the first step of the continuous compliance process, whereby the tool executed at least once during each sprint and produced an actionable list of potential GDPR exposures. However, DataCorp did not regularly add tasks based on the created list of potential GDPR exposures. From our experience, there may be two interlinked causes to the inadequate adoption of continuous compliance in DataCorp. First, employees are constantly busy and finding time to translate GDPR tool results into tasks and subsequently working on such tasks is overly time consuming. This reason is supported by DataCorp's continuous reassurance that

our GDPR tool's results should have been added to the organization's backlog, but employees have been busy with other work. Second, since time is such a valuable resource, employees may feel that the potential GDPR exposures identified by the GDPR tool are not "severe" enough to cause a drastic penalty if temporarily not investigate and resolved. When employees have free time in the future, they could theoretically allocate time to treating and managing the results of our GDPR tool. Moreover, the fact that GDPR continuous compliance has not yet been achieved in DataCorp does not mean that such feat is not possible in the future. As DataCorp hires more employees and matures, it is quite realistic that an employee may be tasked with mending the incomplete continuous compliance cycle and successfully conducting GDPR continuous compliance.

6.1.1 Implications

First, reliance on manual testing is a major challenge to GDPR compliance, but operationalizing GDPR regulations with automated tools is a solution to alleviating manual testing. Additional research into operationalizing other GDPR principles or rights may be a beneficial area of study.

> **Research Implication 1**
>
> How to operationalize and automatically test compliance with the remaining GDPR regulations?

Through operationalization of GDPR regulations, manual testing of GDPR compliance should be replaced with automated verification without the risk of human error or wasting valuable human developer time. This also ties in with organizational shifts to more continuous and automated software development [41]. However, given the difficulty in producing a tool that covers every aspect of privacy, an organization is best served by a complement of tools that together help support an organization's effective treatment of privacy. Continuous compliance is an excellent strategy for GDPR compliance, but buy-in from employees is required, and adhering to continuous compliance steps becomes habitual. CI can help an organization quickly respond to an issue, privacy included. However, an organization may only realize these benefits if potential GDPR exposures are translated into work tasks and these work tasks are later prioritized and resolved.

Practitioner Implication 1

Operationalize GDPR principles into relevant privacy requirements and use automated tests to continuously verify these requirements.

6.2 Insufficient knowledge management impedes privacy awareness and compliance

Treating privacy requirements, such as complying with the GDPR in theory requires employees to have a reasonable level of understanding of these privacy requirements. However, DataCorp did not have a systematic employee training on the GDPR or any specific GDPR regulation. In consequence, an employee had to conduct individual research on the GDPR and knowledge was disseminated on an ad-hoc basis. Furthermore, an employee who was not assigned privacy related tasks nor took own initiative to learn about the GDPR was unlikely to have substantial awareness of the GDPR. For example, months after the GDPR deadline, some employees did not even realize that the GDPR existed. Given the lack of a standard of privacy knowledge that an employee must exhibit, an organization's overall ability to treat privacy was limited. Raising privacy awareness in resource constrained organizations, especially in early development is difficult [11]. Our observation shows that privacy awareness is not only challenging early on, but also difficult in the present and long term. Although it may seem insignificant if even one employee is not aware of the GDPR or its implications, one unintentional accident by an employee can have severe consequences. As demonstrated by cases like the Capital One data breach, one minor configuration mishap in a system's infrastructure can result in the exploitation of millions of users' sensitive data [31].

Moreover, ensuring that an organization's privacy processes is transparent for all employee is also vital to the organization's sufficient treatment of privacy. A lack of shared understanding at DataCorp was exemplified where employees did not always have a strong understanding of privacy processes upstream or downstream of their own work. For example, one employee who works in data processing automatically assumes that what ever his or her is by default GDPR compliant because the data collected is already GDPR compliant. The employee believed that the data handed to the data processing employee were already "treated", which means the employee does

not need to worry or consider the GDPR. Yet, the employee was unsure about any specific privacy safeguards upstream in the process. To effectively manage privacy, automatically assuming GDPR compliance seems inadequate. When asked about the privacy safeguards that exist in upstream processes, the employee was unsure. It is possible that the data processing employee was correct in his or her sentiment, but the employee could have made a more accurate determination if knowledge about each upstream and downstream process was better managed. If transparency of workflow and greater shared understanding existed, the aforementioned scenario would likely be avoided.

As previously stated, complying with the GDPR requires effort from each individual in an organization. For long term compliance, not only is some level of training regarding privacy necessary, but also frequent reiteration of knowledge and awareness to ensure sufficient long term privacy awareness. Standardizing privacy training for employees will help an organization ensure that its employees have a relatively equal background of knowledge to handle privacy requirements. Furthermore, breaking down barriers and increasing shared understanding between employees is paramount to increasing transparency. An employee should be aware of privacy safeguards, not only holistically across an organization, but also in relation to one's own work.

6.2.1 Implications

Insufficient awareness and knowledge management may impede long term compliance because a developer is unlikely able to address a potential privacy problem if the developer is unaware of the pertaining privacy regulation. Therefore, more research is needed to find more efficient strategies to disseminate the GDPR's implications to improve an organization's ability to handle compliance.

Research Implication 2

How to efficiently and effectively disseminate and manage GDPR knowledge?

For effective GDPR compliance, an organization needs to ensure that each employee is adequately trained and aware of the GDPR. An employee must be cognizant of a regulation to consider the regulation during work. Processes and privacy safeguards also must be transparent so that an employee has a breadth of knowledge of the safeguards across an organization.

Practitioner Implication 2

Each employee must be adequately trained about the GDPR. Each employee should be provided role relevant training on GDPR regulations and organizational policies and processes.

6.3 Managers and developers have sharply different priorities for GDPR compliance

DataCorp had a relatively sizeable difference in mentality with respect to the importance and relevance of privacy between managers and developers. In particular, managers value privacy more than developers. This observation was shown by our first survey where managers ranked privacy as the most important NFR to DataCorp's business whereas developers ranked privacy sixth. A second round of surveys occurred seven months after the first survey. In the second round, managers continued to rank privacy as the most important NFR for DataCorp's business, whereas developer sentiment towards privacy experienced only a mild increase if at all. Specifically, privacy dropped to eighth for developers when considering NFR importance for individual work, which we believe is symbolic in the decrease in privacy related work after the GDPR deadline, as a developer had less interactions and personal connection to the GDPR and GDPR compliance. On the other hand, developers' ranking of privacy increased to third from sixth when considering privacy's importance to the organization's business. Despite months after the GDPR deadline, when asked the same question, managers continued to rank privacy as the most important NFR for our collaborating organization. Despite the existence of an external compliance audit, and the presence of external experts, it appears that developers experienced minimal effect on their daily work.

Moreover, regarding GDPR compliance, the fact that privacy tasks were not equally distributed likely contributed to the inequality in valuing privacy. In general, managers were allocated more privacy related tasks than developers. A developer on an individual basis only occasionally if at all received privacy tasks post GDPR deadline. As explained by P8 when asked if he encountered privacy work and/or experienced privacy challenges in work months after our initial interviews, *"No, nothing on the radar"*. Similarly, P7 had a similar response to the same question: *"Not really.*

I never deal with that stuff [privacy]. I am part of the [redacted] team so [privacy is not my concern]". While privacy may be a "nice to have" quality of a feature, privacy does not directly affect a developer work as much as reliability. A developer is unlikely going to be able to release a feature if the new feature degrades the reliability of the overall software. Like reliability, maintainability is a more perturbing quality attribute on a daily basis as code reviews ensure that newly developed code adheres to a standard of quality, whereas privacy can be an NFR that is always desirable, but unlikely to affect the short term development nor release of a feature. Hence, developers may feel less connection to privacy, which may hurt the long term commitment to compliance.

The imbalance in valuing privacy between developers and management reduces the quality of an organization's GDPR compliance, especially long term compliance. As time passes, fewer employees become involved with privacy work, which entails lower awareness and familiarity of privacy. For the purposes of effectively treating privacy using strategies including privacy by design (PbD) [26], developers play a pivotal role [50]. While consultants and lawyers may help interpret and provide guidance on GDPR regulations, developers are ultimately assigned to convert these regulations into requirements and realize the requirements in software. Without developer involvement, an organization's compliance effort is futile. Developers need to have a strong perception and understanding of privacy, but developers often relegate privacy measures to policy based solutions as opposed to architectural changes [50]. Furthermore, organizational discouragement was previously found to be a significant barrier to motivating developers towards privacy [50]. In contrast, DataCorp's managers support adoption of privacy initiatives and prioritize privacy. Ultimately, if managers are highly motivated and intend to promote GDPR compliance within an organization, this motivation is to no avail if people who help realize compliance do not share the same inspiration. To advance towards long term GDPR compliance, an emphasis on privacy must be demonstrated by both managers and developers (i.e. from the "grassroots" level). P1 stated, *"For the entire organization to be GDPR compliant, each individual has to be GDPR compliant and review needs to be done on an individual basis rather than just an organizational perspective"*.

6.3.1 Implications

The difference in mentality between developers and managers towards privacy may be a long term challenge to compliance. Breaking down the barriers between developers and managers and increasing the shared understanding of privacy may be a judicious research exploration.

> **Research Implication 3**
>
> How to break down barriers and increase shared understanding between developers and managers?

To ensure that compliance work is effectively implemented and employees carefully treat privacy, the silo between manager and developer roles must be reduced. As developers ultimately carry out many privacy tasks and interact with data on an everyday basis, a developer must share the same level of urgency and value towards privacy.

> **Practitioner Implication 3**
>
> Increased shared understanding and breaking down "silos" between roles is necessary to ensure that developers have as strong of a concern for privacy as managers.

6.4 Overconfidence in GDPR readiness reduces the visibility of the state of compliance

Although many DataCorp employees had little interaction and felt minimal impact from the GDPR, we were surprised by the confidence exhibited by DataCorp employees. In particular, we were shocked that a large number of employees believed that there are absolutely no privacy exposures within the organization. As stated by P4, *"We are GDPR compliant....We are compliant because we took a proactive approach [to compliance]"*. When asked if there are any privacy exposures in the organization, P8 replied, *"No, I don't feel [there are] any issues"*. While P5 was not as adamant as P4 and P8, P5 still believe that the number of issues was close to minimal. Finally, P6 conceded that no software is "perfect", and that the software probably has privacy exposures, but not many due to the organization's previous effort.

We were surprised by these answers because these answers contrasted with our observation. When we talked with DataCorp's employees, the employees discussed examples of GDPR compliance challenges experienced during their compliance related work. The challenges mentioned by these employees were also reinforced through our observations at the organization. For instance, we observed the difficulty in comprehending some GDPR regulations and the amount of research required to learn more about the GDPR. Provided that the employees who we conversed in our study told us challenges that they indeed experienced and our observations were an accurate representation of the organization's context, we can only conclude that employees were maybe overconfident or had a misrepresented sense of the organization's compliance status.

Furthermore, our observations indicate that developers often did not deal with compliance tasks, especially after the GDPR deadline. Yet, we found that many people who do not spend time on privacy related tasks nor research compliance related material have a firm belief of their knowledge in privacy and the GDPR. In contrast, some employees who deal more with compliance work actually felt that their personal level of GDPR understanding is low. In one particular case, one employee told us that his personal understanding of the GDPR is high because he received some explanations and information about the GDPR from another employee, whom actually had less confidence in their GDPR knowledge. It is possible that the other employee was modest in his self assessment, but the instance exemplifies that some employees are quite confident, perhaps even overly confident in their GDPR understanding and the organization's compliance condition.

6.4.1 Implications

Post GDPR deadline, it is possible that developers spend less time on privacy related tasks. In consequence, a developer may not have the time nor knowledge to comprehensively investigate the system for privacy problems. Blatant confidence may cause oversight or confusion on the true state of compliance of a system. Hence, more research is necessary into improving the transparency and visibility of an organization's GDPR compliance.

Research Implication 4
How can we increase transparency of state of GDPR compliance?

Though having confidence and belief in one's organization may be an important aspect of team bonding and maintaining a strong cultural cohesion, too much confidence may prevent employees from grasping an accurate awareness on weaknesses in the organization's software. Moreover, the false sense of compliance may reduce the willingness to investigate GDPR compliance in more detail. In addition, the false sense may prevent recognition of potential GDPR exposures in an organization's software. As employees are constantly busy as seen in our observations and typical in a startup organization, there may be a sense that compliance is "good enough" and effort can be directed towards more tangible endeavors. An organization's state of compliance needs to be transparent for all employees so that everyone is aware of software elements where compliance is weak.

Practitioner Implication 4

Confident employees is desired, but an organization should ensure that processes and responsibilities are transparent to not allow confidence cloud judgement.

6.5 Offloading privacy concerns relinquishes compliance control to others

DataCorp strongly relies on third party services such as AWS, Azure, and Google Cloud Platform. Due to the reputation of these services, DataCorp believes that it is reasonably protected by these services. First, services like AWS provide state of the art cloud infrastructure, which should have top of the line GDPR compliance strategies and protections that would help protect our collaborating organization's data. Second, these services are in a unique position where they act as the de facto "processors" of our collaborating organization's data. In other words, these processors share the same responsibilities as our collaborating organization to store and process the data in a GDPR compliant manner. Any potential criticism of our collaborating organization would almost certainly apply to the third party service as well. As these third party services are part of particularly large organizations that presumably are GDPR compliant, our collaborating organization seemingly has the aforementioned protections as long as it continues to use these services. However, even industry leading third party services may have vulnerabilities. Third party services may not

be fully GDPR compliant and organizations typically still need to self manage the configuration of its cloud infrastructure.

Large multinational consulting firms offer compliance services where they can send in a team of privacy consultants who review an organization's current state of compliance and provide suggestions for areas to improve. DataCorp hired external help as part of its compliance strategy. As described by P4, [Reputable consultant] reviewed our compliance and he was impressed...[hence, no need to do anything else]. However, solely relying on the positive review of a consultant may provide a false sense of assurance and decrease motivation to further improve the organization's compliance. It is possible that the consultant may have missed an aspect in the review, not to mention, the GDPR may be purposefully ambiguous and consultants often provide differing opinions on the same issue [37]. Given the reputation of the privacy consulting team's parent firm, the belief is that the GDPR compliance certificate provided by the team is reputable; hence, meaningful and potentially perceived as a positive indicator of compliance for our collaborating organization. After conducting a compliance review, DataCorp received a certification of compliance from the consulting team. However, it should be noted that the compliance review is subjective and there is not a standard method or framework to conduct a GDPR compliance review.

Yet, over reliance on these partners to take the brunt of any potential scrutiny may over expose our collaborating organization to external forces. Instead of having full control over its GDPR compliance, the organization may be unintentionally over exposing itself to the compliance of its third party providers and trust that support will come from its customers.

6.5.1 Implications

Partnering with large entities and working with reputable consultants may facilitate offloading or at least sharing privacy responsibilities with these external entities. However, offloading privacy responsibility may result in losing control to external entities. Research is needed into the risks involved with offloading privacy responsibilities and also managing these risks.

Research Implication 5

How can we measure and manage the risks of offloading privacy responsibilities?

An organization can seemingly offload aspects of privacy responsibility to external entities. For example, when an organization uses cloud infrastructure from a third party service, the organization should inherently receive some privacy protections for its data. However, offloading may be a cause for concern because the organization must review whether or not the third party service is sufficiently treating privacy.

Practitioner Implication 5

An organization may offload privacy responsibilities to external entities, but the organization must be cognizant that offloading has risks.

Chapter 7

Threats to Validity

As no research is without limitations, there are a few limitations to our research.

We are transparent with our research approach and methods. We included our interview guide in the appendices as supporting information. However, due to confidentiality reasons, we cannot explicitly describe our collaborator's name nor details. Similarly, we cannot disclose identifying information about our study participants. Nonetheless, we tried to ensure that we had a wide representation of participants, by observing and interviewing employees from a multitude of roles. However, the interview data we collected from our participants may be limited due to participants answering a question a certain way because they know they are being watch (the observer effect). Similarly, our observation and discussion data may be limited as participants may act differently knowing that they were watched. However, we mitigated this limitation by signing participant consent forms and NDAs that explicitly explained that we would disclose any compromising information from any participant. We also reminded participants that we were not at the company to judge or find blame, our research goal was identifying problems and rectifying these problems.

The interpretation of research results may be biased as one co-author has extensive knowledge about DataCorp. However, the extensive knowledge merely served to provide context about DataCorp, not to bias any inferences or conclusions of results.

Reproducibility of the study may be limited as we conducted an in-depth study at one organization and identified and helped resolve specific problems for the one organization. Nonetheless, we expect a study that followed our methodology and techniques in a similar type of organization to find similar challenges and produce similar artifacts.

Lastly, the generalizability of the paper may be limited as we studied only a sin-

gle company. The single company is a small organization (several dozen employees) and operates in the data business, with a reliance on cloud infrastructure. Hence, although we may not be able to generalize our study to other settings such as a large organization developing safety critical software, we expect organizations of similar size and context (i.e. CI practicing, cloud based software, GDPR relevant) to encounter similar challenges as our collaborator. Despite our research focusing on one organization, we did our due diligence strengthening the credibility of our research by conducting an in-depth close collaborative study, with multiple iterations of our operationalized requirements and GDPR tool.

Chapter 8

Conclusion and Future Work

The practices and challenges of GDPR compliance in small organizations practicing CI is still a relatively unexplored area of research. Through design science research, we investigated the compliance challenges in a small, startup organization and identified three primary challenges. We focused on alleviating the challenge of relying on manual GDPR tests and produced two design science artifacts: operationalized GDPR requirements and an automated GDPR tool.

In summary, we developed five implications for research and five implications for practice. Researchers may consider further operationalizing and automatically testing compliance of other GDPR regulations. Researchers may also investigate quantifying the risks of offloading privacy responsibilities and finding a repeatable way to make decisions based on this analysis. Similarly, an organization can operationalize GDPR principles into requirements and continuously test these requirements. Finally, an organization can offload privacy to third parties with the caveat that the organization may lose some control of privacy. We hope our research motivates additional research into GDPR compliance, specifically reducing the challenges associated with GDPR compliance.

Appendix A

Interview Questions Template

Table A.1: Interview Questions Template

Questions
• How has the implementation of the GDPR affected your work? • What do you feel is the most challenging aspect of becoming GDPR compliant? And why? • If you are developing x and developing y. X has a GDPR requirement and Y doesn't, what would you do differently? • What role does privacy requirements play in the organizations sprints or feature development? • How does the organization verify privacy requirements? • How does the organization test privacy requirements? • How does the organization verify privacy requirements • Do you have a good idea on the number of privacy issues currently within the organization? • How long would it take for the organization to manually find all GDPR violations in its system. How about a specific area environment (e.g. AWS)? • Has the CI Process helped with GDPR adoption? If so, in what ways? • Are you aware of any new privacy regulations?

Appendix B

GDPR Tool Scan Results

Table B.1: Number of Infrastructure Resources Scanned by GDPR Tool (W represents Week, For confidentiality purposes, totals are rounded to the nearest 25 and anything below 25 is rounded to 25)

	1st Iteration				2nd Iteration				3rd Iteration				4th Iteration			
Infrastructure Type	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**	**W11**	**W12**	**W13**	**W14**	**W15**	**W16**
Database	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
Server	150	150	150	150	175	175	150	150	150	150	150	150	175	175	200	175
Route Gate	50	50	50	25	50	50	50	50	50	50	50	50	50	50	50	50
Load Balancer Type 1	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
Load Balancer Type 2					25	25	25	25	25	25	25	25	25	25	25	25
Router	100	100	100	100	100	100	100	100	100	100	100	100	75	75	75	75
Cloud Storage									125	125	125	125	125	125	125	125
Cloud Firewall	175	175	175	175	175	175	175	175	175	175	175	175	175	175	175	175
Server Storage	175	175	175	175	175	175	175	175	175	175	175	175	200	200	200	200
Cloud Network	50	50	50	25	50	50	50	50	50	50	50	50	50	50	50	50
Access Management													200	200	200	200
Rounded Total	**775**	**775**	**775**	**725**	**825**	**825**	**800**	**800**	**925**	**925**	**925**	**925**	**1150**	**1150**	**1175**	**1150**

Table B.2: Number of Potential GDPR Exposures Identified by GDPR Tool per Infrastructure Resource (W represents Week of Scan, Average represents ratio of exposures per unit of resource)

	1st Iteration				2nd Iteration				3rd Iteration				4th Iteration			
Infrastructure Type	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**	**W11**	**W12**	**W13**	**W14**	**W15**	**W16**
Database	2.55	2.55	2.55	2.55	4.43	4.43	4.43	4.43	4.43	4.43	4.43	4.48	4.50	4.50	4.50	4.50
Server	0.22	0.22	0.21	0.18	0.20	0.18	0.18	0.17	0.16	0.18	0.15	0.17	0.19	0.21	0.24	0.24
Route Gate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Load Balancer Type 1	4.55	4.55	4.55	4.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Load Balancer Type 2					0.67	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Router	0.36	0.36	0.36	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cloud Storage									0.57	0.57	0.54	0.54	0.54	0.54	0.53	0.53
Cloud Firewall	2.15	2.16	2.16	2.17	2.17	2.17	2.16	2.16	2.16	2.16	2.16	2.16	2.14	2.14	2.15	2.15
Server Storage	0.76	0.76	0.78	0.78	0.77	0.76	0.78	0.78	0.77	0.77	0.77	0.77	0.58	0.57	0.55	0.53
Cloud Network	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Access Management													0.64	0.64	0.64	0.68
Average Across all Resources	**0.99**	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**0.93**	**0.93**	**0.93**	**0.93**	**0.87**	**0.86**	**0.86**	**0.86**

Table B.3: Number of Potential GDPR Exposures by GDPR Tool per Infrastructure region (W represents Week of scan, For confidentiality purposes, totals are rounded to the nearest 15 and anything below 15 is set to 15))

	1st Iteration				2nd Iteration				3rd Iteration				4th Iteration			
Region	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**	**W11**	**W12**	**W13**	**W14**	**W15**	**W16**
Region-A					15	15	15	15	15	15	15	15	15	15	15	15
Region-B	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-C									75	75	75	75	195	195	195	195
Region-D	30	30	30	30	30	30	30	30	30	30	30	30	15	15	15	15
Region-E	30	30	30	30	30	30	30	30	30	30	30	30	15	15	15	15
Region-F	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-G	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-H	435	435	420	405	450	450	450	435	435	435	435	435	435	450	450	450
Region-I	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-J	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-K	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-L	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-M	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-N	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-O	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Region-P	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
Region-Q													15	15	15	15
Region-R	15	15	15	15	15	15	15	15	15	15	15	15	30	30	30	30
Region-S	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
Rounded Total	**780**	**780**	**765**	**750**	**810**	**810**	**810**	**795**	**870**	**870**	**870**	**870**	**990**	**1005**	**1005**	**1005**

Table B.4: Number of Potential GDPR Exposures Identified by GDPR Tool per GDPR Principle (W represents Week, For confidentiality purposes, totals are rounded to the nearest 50 and anything below 50 is set to 50)

	1st Iteration				2nd Iteration				3rd Iteration				4th Iteration			
GDPR Principle	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**	**W11**	**W12**	**W13**	**W14**	**W15**	**W16**
Integrity and Confid.	650	650	650	650	600	600	550	550	650	650	650	650	750	750	750	750
Storage Minimization	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
Data Minimization	50	50	50	50	150	150	150	150	150	150	150	150	150	150	150	150
Rounded Total	**750**	**750**	**750**	**750**	**800**	**800**	**750**	**750**	**850**	**850**	**850**	**850**	**950**	**950**	**950**	**950**

Table B.5: Number of Potential GDPR Exposures identified by GDPR Tool per GDPR Recital (W represents Week of scan, For confidentiality purposes, totals are rounded to the nearest 50)

	1st Iteration				2nd Iteration				3rd Iteration				4th Iteration			
GDPR Recital	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W9**	**W10**	**W11**	**W12**	**W13**	**W14**	**W15**	**W16**
Recital 83	650	700	700	650	600	600	550	550	650	650	650	650	750	750	750	750
Recital 39	100	100	100	100	150	150	150	150	150	150	150	150	150	150	150	150
Rounded Total	**750**	**800**	**800**	**750**	**750**	**750**	**700**	**700**	**800**	**800**	**800**	**800**	**900**	**900**	**900**	**900**

Appendix C

Publications

Continuous Requirements: An Example Using GDPR

Ze Shi Li
University of Victoria
Victoria, BC, Canada
lize@uvic.ca

Colin Werner
University of Victoria
Victoria, BC, Canada
colinwerner@uvic.ca

Neil Ernst
University of Victoria
Victoria, BC, Canada
nernst@uvic.ca

***Abstract*—Recently, a stringent set of privacy regulations, the General Data Protection Regulation (GDPR), was enacted in the European Union, which can be considered a privacy non-functional requirement (NFR). As a result, an organization that collects or processes data from European citizens must adhere to the GDPR. Previous studies have shown that compliance to the GDPR poses a number of challenges, which we have confirmed in our own research. In this paper, we describe our ongoing research collaboration with a startup organization that is adopting the GDPR. In addition, during the course of our research, we found that our industry collaborator, practices continuous integration (CI) like many other organizations. The number of organizations adopting CI has increased since Fowler first published his definition of CI. As such, another aspect of our current research is exploring the effects of CI on privacy NFRs and other NFRs. Finally, we describe a design science approach to iteratively learn about industry challenges in GDPR compliance, NFRs in the context of CI, as well as our ongoing work creating a tool to potentially mitigate observed GDPR compliance challenges.**

I. Introduction

The General Data Protection Regulation (GDPR) is a European Union (EU) legislation with potentially significant effects on how an organization manages personally identifiable information (PII) for EU citizens that was passed on May 24, 2016 [1]. The GDPR's mandate is to protect the privacy of EU citizens and will punish an organization that fails to comply with the GDPR, regardless of whether the organization did so with intent or negligence. Although the GDPR was not immediately enforced, an organization conducting business in the EU was expected to obtain GDPR compliance by May 25, 2018 [1]. For an organization actively working to attain compliance, the organization may operationalize each GDPR regulation into smaller work tasks. These smaller work tasks that help an organization adhere to the GDPR can be thought of as privacy non-functional requirements (NFRs).

NFRs act as characteristics that may guide the design and implementation of an organization's systems [2]. These characteristics often influence architectural decisions, which may have significant impact on a system. Architectural decisions may affect the answer to questions such as: How modular is a system? Is the system maintainable? How does the system deal with customer data? Accordingly, researchers often describe NFRs as important aspects of a system. Yet, researchers do not have a consensus on a universally accepted definition of an NFR. Moreover, the concept of NFRs is largely neglected in practice [3]. Since customers and stakeholders frequently express desired outcomes of a system in terms of functionality [4], product development focuses on satisfying the stated functions. When product design and development solely concentrate on functionality, as opposed to the underlying NFRs with long term ramifications that impact architecture, an organization may be required to re-factor a considerable amount of code.

For a startup organization that is pursuing growth and shipping features to customers as fast as possible, NFRs are not an initial priority [5]. Once a startup organization is able to hire more employees, the organization may be able to address more NFRs. However, even if a startup organization emphasizes some NFRs, such as privacy, from the onset, external forces may overwhelm an organization's initial preparation. Due to the magnitude and comprehensiveness of the GDPR, a startup organization that believes and practices the notion of protecting user privacy potentially may not have the necessary amount of resources to prepare itself for compliance.

In part due to our own interest in privacy and NFR research, as well as being approached by a local startup organization, which was required to comply with the GDPR, we began this GDPR and NFR research. Our partner organization is based in Canada, but has many European users. Therefore, the organization falls under the jurisdiction of the GDPR. Although our partner organization maintains a firm belief in user privacy, the organization experienced challenges in its initial adoption of the GDPR and desired robust processes, policies, and controls to help protect personal data and comply with the GDPR. The existence of their challenges echos the large number of organizations who reported difficulty complying by the GDPR deadline [6] [7]. To address our partner organization's problems and our research intrigue of the GDPR, we adopted a design science research methodology, which involves iterative research. The nature of iterative research allows us to build and refine our research artifacts until the completion of the research. Additionally, our research methodology contained two envisioned artifacts. The first artifact is the current state of practice of an organization adopting and complying with the GDPR and the challenges that an organization faces during

its compliance, since we did not have a clear idea of the GDPR adoption challenges nor adoption practices adhered by our partner organization. Based on our analysis of the first artifact, the second artifact contains a GDPR tool that may aid an organization's GDPR compliance.

The novelty of our research is to identify *what* a developer did to prepare for the GDPR and *how* the GDPR affects a developer. In particular, learning about industrial practices strengthens the knowledge base of specific challenges that plague an organization's GDPR compliance.

While our research is currently ongoing, we will discuss our preliminary results in this paper, which includes parts of the first artifact. Currently, our iterative research is in the midst of producing the second artifact. When we first began our research we quickly found another intriguing aspect that was ubiquitous to our partner organization's work; although we had not initially considered the effect of continuous integration (CI) [8] as part of our research. We found that CI had a large impact on our partner's planning, development, testing, and deployment. Hence, we could not, and cannot, disregard the importance of CI on our partner organization, especially the impact of CI on the organization's treatment of NFRs and compliance with the GDPR.

We argue that this paper presents three main contributions

- detailed analysis of environmental context for a startup dealing with GDPR, in a CI context,
- methodology for investigating CI and GDPR in practice, and
- preliminary results on developer awareness of privacy and GDPR compliance.

The rest of the paper is structured as follows. In Section II, an overview of NFRs, the GDPR, CI, and related work in privacy is covered. In Section III, we provide details on our research methodology. In Section IV, we share some insights from our initial findings. In Section V, we conclude the paper and highlight our future work.

II. Background and Related Work

A. Non-Functional Requirements

As previously mentioned, the GDPR may be considered as a single or a series of privacy NFRs. A common definition of an NFR is provided by Glinz [2] as "an attribute of or constraint on a system". In essence, an NFR, also known as a quality attribute, is a characteristic of a system that may help guide architectural decisions. In practice, NFRs are often ignored or not prioritized by both customers and developers [9] or prioritized lower than functional requirements [10].

Deciding which NFRs are important to a particular system is a challenging task, as decisions must be made with trade-offs between various NFRs. In addition, the lack of any sort of prioritization is also a form of a trade-off. Whether willingly or not, when an organization settles on an NFR trade-off, the organization may inadvertently accumulate more technical debt. As the amount of technical debt swells, the organization's ability to develop new features may be limited by the insurmountable technical debt. Eventually, such an organization may reach a point where substantial code re-factoring is necessary for the organization to continue to deliver new features. As a result, an organization that neglects NFRs when developing and designing a system may need to significantly re-factor its system, if at all possible, later in the system's life cycle [4]. For example, an NFR, such as privacy, can be especially difficult to implement at a later stage in a product's life cycle and can lead to a "long and unhappy history of incremental patching and retrofitting that characterizes the current Internet architecture" [11]. Furthermore, attempting to retrofit privacy requirements may be flawed without initial consideration of security as security and privacy are often intertwined [12].

As privacy safeguards and procedures are a fundamental aspect of a system, privacy measures should inherently exist as part of the product from inception. On the contrary, an organization may perceive benefit if an appropriate amount of importance is placed on NFRs early in a product life cycle, which can help propel and shape the architectural design and implementation [13].

B. General Data Protection Regulation

Prior to the passing of the GDPR, our partner organization already implemented controls to protect personal data, but as the most comprehensive privacy law, the GDPR affected everything from development to storage of data. In addition to the GDPR, other major privacy regulations have been or will be introduced, including the California Consumer Privacy Act (CCPA),[1] as well as Vermont's Data Broker Regulations.[2] In addition to being a complicated NFR to handle, privacy is one NFR that applies to most organizations as the GDPR applies to any organization that is based in the EU or collects personal data from identifiable EU citizens [1], which certainly is applicable to a significant number of organizations.

As a novel and trailblazing set of regulations that strive to protect the privacy of individual personal data from abuse, the GDPR stipulates tough financial penalties for violators and imposes stringent regulations to an organization's data collection and process practices. Consequently, many organizations have adopted the GDPR as their default privacy standard, as the GDPR is generally regarded as the most strict privacy policy released and enforced to date [14].

However, as the deadline approached, many organizations were still not yet GDPR compliant [6] [7]. Leading up to the 2018 deadline, a non-GDPR compliant organization had two options: 1) turn off aspects of their system that were not GDPR compliant or 2) continue with normal operations in hope of not being caught for non-compliance. Unfortunately, statistics are unavailable on the percentage of organizations that risked being found in contempt of the GDPR and continued with their usual operations despite being non-compliant. Finally, the crux

[1]https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[2]https://legislature.vermont.gov/Documents/2018/Docs/ACTS/ACT171/ACT171%20As%20Enacted.pdf

of the GDPR is the emphasis on being compliant *all* of the time. Instead of having the luxury to have discretion on when to comply with the GDPR, our partner organization must be continually compliant.

C. Continuous Integration

To better help our partner organization, we had to first understand their development process. However, we not only considered their development process, but also planning, testing, and deployment processes. Our partner organization is a proponent of CI. CI is a practice that involves the use of a.) automated builds and tests, b.) each developer committing at least once a day, and c.) quickly fixing broken builds [8]. A characteristic of an organization that practices CI is that the organization quickly releases new updates to users and rapidly receives feedback from users [8].

In the current competitive business environment, it is difficult for an organization to compete if the organization is unable to quickly adapt to changes in the market and release products that match customer demands. As such, an organization may opt to adopt CI. The roots of CI originated from Agile development, Lean development, and Extreme Programming [8] and has extended to other "continuous" practices such as continuous delivery (CD) [15]. The appeal for an organization to adopt continuous practices, includes fast release of software to customers, reliably releasing to production, and high customer satisfaction [16].

CI aims to reduce a developer's time spent on manual tasks, such as building and testing software [8]. Similarly, an organization practicing CD not only practices CI, but also maintains a high-level of confidence that the organization's software is in a production-ready state at all times [15]). CD aims to maintain a high-level of code quality and also enables an organization to shorten a customer's wait time for receiving the latest updates and remove the hardships of preparing software for release [17].

Unfortunately, there have been conflicting reports on operationalizing NFRs in the context of "continuous". One study has suggested it is difficult to automate testing of NFRs [17] as creating validation criteria for NFRs is not clear-cut. Practitioners in a recent summit indicate that NFRs, such as privacy, are not the concern of every developer [18]. Nevertheless, if an organization is serious about GDPR compliance, a privacy NFR, then the organization may require a privacy-conscience effort from every employee.

Furthermore, considering that the GDPR expects continuous compliance, it may be prudent for an organization's continuous pipeline to run a series of GDPR-related tests on every commit to ensure that privacy is not compromised. However, we do not know the ideal frequency at which an organization should test privacy. Moreover, a frequency that is optimal for one organization may not translate to another organization. A goal of our research is to help contextualize the frequency and depth of testing the practitioner's perspective. If an organization forgoes the notion of NFRs and develops software without consideration of constraints or attributes of a system, the organization may accumulate substantial amounts of technical debt [19]. When an organization accumulates a crushing amount of technical debt, the software may experience catastrophic challenges, such as the inability to re-factor architecture and loss of usability [20].

D. Related Privacy Work

Seminal works in privacy include Deng et al's [21] LINDDUN methodology that can help identify privacy threats in an organization's system and map those threats to privacy requirements, but the methodology abstracts threats at a high level. Nonetheless, just as Deng et al. suggested solution strategies for privacy requirements [21], our partner organization shut off parts of their EU services before the implementation of the GDPR. While previous studies on helping organization assess GDPR compliance exist, such as a tool based approach to conduct Data Protection Impact Assessments [22], there is a shortage of studies analyzing GDPR adoption *challenges* and *practices*. In Sirur et al.'s [23] work, it was found that large organizations did not experience significant obstacles with GDPR compliance. Alternatively, smaller organizations without significant prior emphasis on security felt GDPR compliance was onerous [23] indicating there was insufficient privacy by design [24]. Furthermore, the surveys leading up to the compliance deadline [6] [7] indicated that a substantial number of organizations were not ready for the GDPR in time. Regardless, there appears to be a major disconnect between privacy concerns, such as the GDPR, and implementing engineering solutions to satisfy privacy concerns [25] [26]. Methodologies do exist that focus on engineering with privacy concerns, such as privacy by design; however, privacy by design revolves around the ability to develop a system from the onset, as opposed to retrofitting privacy into an existing, perhaps legacy, system [27]. An organization adopts, often due to convenience, the less reliable "privacy-by-policy" approach as opposed to the more reliable "privacy-by-architecture" [28]; however, "privacy-by-architecture" is difficult to apply to an existing project, much like "privacy-by-design", as re-factoring a business model may be even more challenging than solely a system.

Ultimately, regardless of how a system is designed or built, there is a need to ensure that the system remains continuously compliant with the GDPR mandate. One study proposes reusing static code analysis tools to define and discover potential GDPR compliance violations [29]. Unfortunately, static code analysis is limited to analyzing only code, whereas GDPR compliance encompasses many more components, such as infrastructure, architecture, and databases. Another proprietary solution, IBM Security Guardium Analyzer, is able to analyze a database and classify a datum as personally identifiable information [30]. Hewlett Packard Enterprises offers a similar data classification tool as part of their GDPR Starter Kit[3].

[3]https://www.hpe.com/us/en/newsroom/news-advisory/2017/05/hpe-software-launches-gdpr-starter-kit-to-expedite-and-simplify-compliance.html

However, simply raising awareness by identifying potentially personally identifiable information is only one, albeit quite large, aspect of GDPR compliance. This lack of an all-encompassing continuous GDPR compliance tool motivates our research ambition to collaborate with industrial partners in a design science methodology to help design, implement, and assess a continuous GDPR compliance tool.

III. Design Science Methodology

The purpose of our study is to increase the understanding of the GDPR and privacy NFRs in an industrial setting. In particular, we observe and analyze how an organization practicing CI manages and handles NFRs. As previously discussed, the current state of knowledge of an organization practicing continuous practices suggests that the organization may not clearly define or test NFRs. Therefore, one of the main goals of our research is studying how an organization deals with privacy NFRs that the organization must prioritize and comply. During our research, we are documenting GDPR adoption challenges of our partner organization to increase the awareness of adoption pitfalls. More importantly, we want to help alleviate some of these challenges; as such, we are designing and developing a GDPR compliance tool to mitigate some GDPR compliance challenges and ensuring the tool is felicitous for practitioners. The nature of our research started with the exploration of NFRs (specifically privacy) during our partner's adoption and compliance with the GDPR. Since we began the research without knowing the specific challenges and practices of our collaborating organization, we are conducting design science research, which relies on iterative cycles of building and refining our design science artifacts [31].

Our research methodology consists of two major parts with a total of four design science iterations as shown in Fig. 1. *Part A* involved interviewing developers from our collaborating organization and contained one design science iteration. *Part B* involves designing, developing, deploying, and synthesizing the results of a GDPR tool that aids our collaborating organization. To build a tool that benefits our collaborating organization, we observe and reflect on each iteration from both parts *A* and *B*. The first iterative cycle was a problem investigation cycle, which forms the foundation for, and informs the work of, our subsequent cycles. The primary focus of the first cycle was to identify the standard of practice of defining and testing NFRs observed in practice and explained by our collaborating organization. Furthermore, we interviewed our collaborating organization to gain insight on their GDPR adoption practices and challenges as a specific example of an NFR in practice. Finally, we conducted our research mindful of CI. We observed our collaborating organization's development process and also included questions about CI in our interviews. In short, the design science artifact produced from the first cycle is the breadth of knowledge that represents an organization's definition and treatment of NFRs, as well as challenges experienced by practitioners when complying with the GDPR.

Currently, we are past the first iteration cycle and are working on *Part B* as shown in Fig. 1. Although the relevance cycle of our research has been established, the rigor cycle and evaluation of our research artifacts are still ongoing. Based on the insights gained from the first iteration, we are iteratively developing, and deploying a GDPR tool as part of the second, third, and fourth iterations. After each deployment of the GDPR tool, we further observe and interview developers to understand the positive impact of the tool; any suggestions for improvement or observed deficiencies are documented and congregated into the subsequent iteration. Similar to why our partner organization practices CI, we can quickly adapt our tool based on feedback after deployment. As a result, the third and fourth iterations update the GDPR tool by reflecting on the tool through the tool's execution. Thus, we are producing two artifacts, 1) the knowledge-base built during our initial iteration and 2) a tool that helps assess the understanding of our knowledge-base in a practical industry setting.

The interviews and observations from *Part A* contain various themes that represent practices and challenges. To reduce misinterpretations, our primary form of evaluation is performed by validating our interpretation with our collaborating organization. As part of our ongoing work regarding *Part B*, we are creating a GDPR tool that is being deployed in each design cycle iteration. As a result, the GDPR tool has three versions, each derived and built upon its predecessor. After deploying each version, our collaborating organization analyzes the results of the tool and submits feedback regarding the strengths and weaknesses of the tool. Similarly, we observe the results of the tool to form potential ideas for future enhancements. Our tool is evolving through iterative deployment and evaluation, the result of which enables us to produce an artifact that is both relevant to industrial practices and novel to the current knowledge base.

IV. Preliminary Results

As previously described, a major aspect of our first iterative cycle involved observing and interviewing our collaborating organization. In our initial analysis, our data suggests that the scope of the GDPR was far greater than expected and represented an immense undertaking for a startup organization. A large organization may be able to designate a group of employees to work predominantly on GDPR compliance work, but our collaborating organization (less than 50 employees) was restricted by the amount of resources it could allocate towards GDPR compliance. Furthermore, we noted in our observations and interviews that a developer was usually busy with various tasks, which further complicated understanding the GDPR, especially given the size of the GDPR. The benefit of CI towards facilitating rapid releases and quick feedback seemed to come at the cost of the thorough definition and characterization of an NFR. Notwithstanding a developer's exuberant effort, a developer is not likely to be a legal expert nor trained in privacy law. Hence, the constant affair of balancing different tasks and lack of GDPR training impeded our partner organization's ability to sufficiently comply. Moreover, upon

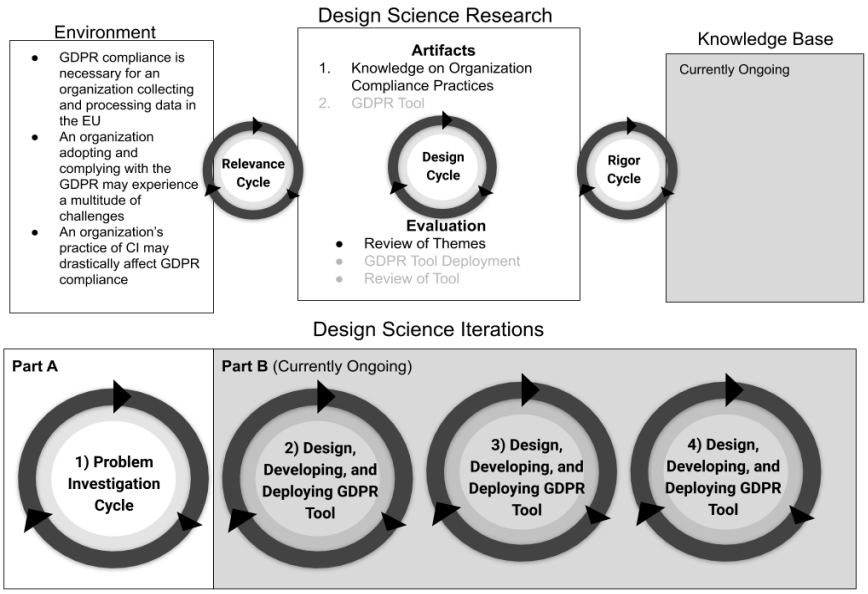

Fig. 1: Design Science Methodology

Note: The greyed parts of the Design Science Research diagram represent ongoing work. The Design Science Iterations diagram illustrates the four cycles of our research that will help produce the two artifacts

close investigation, we found that a developer typically had little influence or relationship to privacy in his or her work, yet the adoption and compliance to the GDPR did affect the same developer's work. This led to a disconnect between a developer's previous work and the developer's ideal GDPR compliant work. Although change was minimal for the vast majority of developers in our collaborating organization, the fact remained that almost every developer had to conduct some level of research for the GDPR. Furthermore, most developer agreed that finding a GDPR exposure in their system is difficult when considering the size of their system and the GDPR.

Our results shed light on the fact that challenges do exist in our collaborating organization's GDPR compliance, but also provide specific challenges that we can help tackle. Knowing that time is of essence for developers, it may be possible for our GDPR tool to help flag GDPR exposures so that a developer does spend an exorbitant amount of time researching and finding specific GDPR exposures in their system.

V. Conclusions

In this paper, we highlight our current study on how an organization deals with the GDPR, NFRs, and CI in practice. In particular, we highlight the need to further study those topics due to the enactment of the GDPR and increasingly widespread use of CI. We also proposed and discussed a design science approach that we currently conduct with a collaborating organization that practices CI. Furthermore, we provide details on some of our observations and interview results from our initial work. As described in Section III and Fig. 1, our approach has four iterative cycles split into two parts *A* and *B*. *Part A* had one iterative cycle and involved identifying how practitioners handle and comply with the GDPR. The first iterative cycle not only included studying practitioner practices, but also learning and identifying challenges experienced by practitioners when complying with the GDPR. In contrast, *Part B* relies on the findings in *Part A* and has three iterative cycles. The purpose of *Part B* is to iteratively design and build a GDPR that assuages some compliance challenges identified in *Part A*.

The current state of knowledge on the relationship between NFRs and CI indicates little NFR prioritization and definition. Yet, we are auspicious that our research will help bridge the knowledge gap between the GDPR, NFRs, and CI. We believe the discovered GDPR compliance challenges and our potential GDPR tool will be beneficial contributions to the research and industry community. Nonetheless, more work is necessary to investigate our initial findings.

References

[1] *Data protection in the EU*. en. Text. URL: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en (visited on 04/06/2019).

[2] M. Glinz. "On Non-Functional Requirements". In: *15th IEEE International Requirements Engineering Conference (RE 2007)*. Oct. 2007, pp. 21–26. DOI: 10.1109/RE.2007.45.

[3] J. Eckhardt, A. Vogelsang, and D. M. Fernández. "Are "Non-functional" Requirements really Non-functional? An Investigation of Non-functional Requirements in Practice". In: *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. 2016, pp. 832–842. DOI: 10.1145/2884781.2884788.

[4] L. Cao and B. Ramesh. "Agile Requirements Engineering Practices: An Empirical Study". In: *IEEE Software* 25.1 (Jan. 2008), pp. 60–67. ISSN: 0740-7459. DOI: 10.1109/MS.2008.1.

[5] C. Gralha, D. Damian, A. I. T. Wasserman, M. Goulão, and J. Araújo. "The Evolution of Requirements Practices in Software Startups". In: *Proceedings of the 40th International Conference on Software Engineering*. ICSE '18. event-place: Gothenburg, Sweden. New York, NY, USA: ACM, 2018, pp. 823–833. ISBN: 978-1-4503-5638-1. DOI: 10.1145/3180155.3180158. URL: http://doi.acm.org/10.1145/3180155.3180158 (visited on 07/09/2019).

[6] C. F.i.S. o. May 16, 2018, and. A. Pst. *Only 36% of firms will be fully compliant with GDPR by its deadline*. en. URL: https://www.techrepublic.com/article/only-36-of-firms-will-be-fully-compliant-with-gdpr-by-its-deadline/ (visited on 04/06/2019).

[7] H. N. Security. *Only 20% of companies have fully completed their GDPR implementations*. en-US. July 2018. URL: https://www.helpnetsecurity.com/2018/07/16/complete-gdpr-implementation/ (visited on 04/06/2019).

[8] M. Fowler. *Continuous Integration*. URL: https://martinfowler.com/articles/continuousIntegration.html.

[9] F. Buschmann, D. Ameller, C. P. Ayala, J. Cabot, and X. Franch. "Architecture Quality Revisited". In: *IEEE Software* 29.4 (2012), pp. 22–24. ISSN: 0740-7459. DOI: 10.1109/MS.2012.77.

[10] L. Chung and B. A. Nixon. "Dealing with Non-functional Requirements: Three Experimental Studies of a Process-oriented Approach". In: *Proceedings of the 17th International Conference on Software Engineering*. ICSE '95. Seattle, Washington, USA: ACM, 1995, pp. 25–37. ISBN: 0-89791-708-1. DOI: 10.1145/225014.225017. URL: http://doi.acm.org/10.1145/225014.225017.

[11] M. Ambrosin, A. Compagno, M. Conti, C. Ghali, and G. Tsudik. "Security and Privacy Analysis of National Science Foundation Future Internet Architectures". In: *IEEE Communications Surveys Tutorials* 20.2 (2018), pp. 1418–1442. ISSN: 1553-877X. DOI: 10.1109/COMST.2018.2798280.

[12] G. Kalogridis, M. Sooriyabandara, Z. Fan, and M. A. Mustafa. "Toward Unified Security and Privacy Protection for Smart Meter Networks". In: *IEEE Systems Journal* 8.2 (2014), pp. 641–654. ISSN: 1932-8184. DOI: 10.1109/JSYST.2013.2260940.

[13] L. Chung and J. C. S. do Prado Leite. "On Non-Functional Requirements in Software Engineering". In: *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*. Ed. by A. T. Borgida, V. K. Chaudhri, P. Giorgini, and E. S. Yu. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 363–379. ISBN: 978-3-642-02463-4. DOI: 10.1007/978-3-642-02463-4_19. URL: https://doi.org/10.1007/978-3-642-02463-4_19.

[14] *What is GDPR, the EU's new data protection law?* 2019. URL: https://gdpr.eu/what-is-gdpr/.

[15] L. Chen. "Continuous Delivery: Overcoming adoption challenges". en. In: *Journal of Systems and Software* 128 (June 2017), pp. 72–86. ISSN: 01641212. DOI: 10.1016/j.jss.2017.02.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0164121217300353 (visited on 01/14/2019).

[16] L. Chen. "Towards Architecting for Continuous Delivery". In: *2015 12th Working IEEE/IFIP Conference on Software Architecture*. May 2015, pp. 131–134. DOI: 10.1109/WICSA.2015.23.

[17] L. Chen. "Continuous Delivery: Overcoming adoption challenges". en. In: *Journal of Systems and Software* 128 (June 2017), pp. 72–86. ISSN: 01641212. DOI: 10.1016/j.jss.2017.02.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0164121217300353 (visited on 01/14/2019).

[18] C. Parnin, E. Helms, C. Atlee, H. Boughton, M. Ghattas, A. Glover, J. Holman, J. Micco, B. Murphy, T. Savor, M. Stumm, S. Whitaker, and L. Williams. "The Top 10 Adages in Continuous Deployment". In: *IEEE Software* 34.3 (May 2017), pp. 86–95. ISSN: 0740-7459. DOI: 10.1109/MS.2017.86.

[19] N. A. Ernst. "On the Role of Requirements in Understanding and Managing Technical Debt". In: *Proceedings of the Third International Workshop on Managing Technical Debt*. MTD '12. Zurich, Switzerland: IEEE Press, 2012, pp. 61–64. ISBN: 978-1-4673-1749-8. URL: http://dl.acm.org/citation.cfm?id=2666036.2666047.

[20] W. Cunningham. "The WyCash Portfolio Management System". In: *SIGPLAN OOPS Mess.* 4.2 (Dec. 1992), pp. 29–30. ISSN: 1055-6400. DOI: 10.1145/157710.157715. URL: http://doi.acm.org/10.1145/157710.157715.

[21] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. "A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements". en. In: *Requirements Engineering* 16.1 (Mar. 2011), pp. 3–32. ISSN: 1432-010X. DOI: 10.1007/s00766-010-0115-7. URL: https://doi.org/10.1007/s00766-010-0115-7 (visited on 07/26/2019).

[22] J. Coles, S. Faily, and D. Ki-Aries. "Tool-Supporting Data Protection Impact Assessments with CAIRIS". In: *2018 IEEE 5th International Workshop on Evolving Security Privacy Requirements Engineering (ESPRE)*. Aug. 2018, pp. 21–27. DOI: 10.1109/ESPRE.2018.00010.

[23] S. Sirur, J. R. C. Nurse, and H. Webb. "Are we there yet? Understanding the challenges faced in complying with the General Data Protection Regulation (GDPR)". In: *arXiv:1808.07338 [cs]* (Aug. 2018). arXiv: 1808.07338. URL: http://arxiv.org/abs/1808.07338 (visited on 04/06/2019).

[24] A. Cavoukian. "Privacy by design: the definitive workshop. A foreword by Ann Cavoukian, Ph. D". In: *Identity in the Information Society* 3.2 (2010), pp. 247–251.

[25] Y. Martin and A. Kung. "Methods and Tools for GDPR Compliance Through Privacy and Data Protection Engineering". In: *2018 IEEE European Symposium on Security and Privacy Workshops (EuroS PW)*. 2018, pp. 108–111. DOI: 10.1109/EuroSPW.2018.00021.

[26] A. Senarath and N. A. G. Arachchilage. "Why Developers Cannot Embed Privacy into Software Systems?: An Empirical Investigation". In: *Proceedings of the 22Nd International Conference on Evaluation and Assessment in Software Engineering 2018*. EASE'18. Christchurch, New Zealand: ACM, 2018, pp. 211–216. ISBN: 978-1-4503-6403-4. DOI: 10.1145/3210459.3210484. URL: http://doi.acm.org.ezproxy.library.uvic.ca/10.1145/3210459.3210484.

[27] S. Gürses, C. Troncoso, and C. Diaz. "Engineering privacy by design". In: *Computers, Privacy & Data Protection* 14.3 (2011), p. 25.

[28] S. Spiekermann and L. F. Cranor. "Engineering Privacy". In: *IEEE Transactions on Software Engineering* 35.1 (Jan. 2009), pp. 67–82. ISSN: 0098-5589. DOI: 10.1109/TSE.2008.88.

[29] P. Ferrara and F. Spoto. "Static Analysis for GDPR Compliance". In: *ITASEC*. 2018.

[30] D. Shah, L. Lindsay, J. Diaz, S. Shechter, and A. Becher. "IBM Security Guardium Analyzer Bootcamp". In: *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering*. CASCON '18. Markham, Ontario, Canada: IBM Corp., 2018, pp. 380–382. URL: http://dl.acm.org.ezproxy.library.uvic.ca/citation.cfm?id=3291291.3291349.

[31] A. R. Hevner, S. T. March, J. Park, and S. Ram. "Design Science in Information Systems Research". In: *MIS Quarterly* 28.1 (2004), pp. 75–105. ISSN: 02767783. URL: http://www.jstor.org/stable/25148625.

Bibliography

[1] Data protection in the EU. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en.

[2] E.U. privacy regulations cause some web services to block European visitors. https://www.nbcnews.com/tech/tech-news/chicago-tribune-los-angeles-times-block-european-users-due-gdpr-n877591.

[3] The principles. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

[4] What is an SME? | Internal Market, Industry, Entrepreneurship and SMEs. https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en.

[5] 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) (Text with EEA relevance), May 2016.

[6] NPR & GDPR: Users that decline cookies sent to a plain text website. https://ppc.land/npr-gdpr-users-that-decline-cookies-sent-to-a-plain-text-website/, May 2018.

[7] Alaa Altorbaq, Fredrik Blix, and Stina Srman. Data subject rights in the cloud: A grounded study on data protection assurance in the light of GDPR. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 305–310, December 2017.

[8] M. Ambrosin, A. Compagno, M. Conti, C. Ghali, and G. Tsudik. Security and privacy analysis of national science foundation future internet architectures. *IEEE Communications Surveys Tutorials*, 20(2):1418–1442, Secondquarter 2018.

[9] D. Ameller, C. Ayala, J. Cabot, and X. Franch. How do software architects consider non-functional requirements: An exploratory study. In *2012 20th IEEE International Requirements Engineering Conference (RE)*, pages 41–50, September 2012.

[10] Jorge Aranda, Steve Easterbrook, and Greg Wilson. Requirements in the wild: How small companies do it. In *15th IEEE International Requirements Engineering Conference (RE 2007)*, pages 39–48, Delhi, India, October 2007. IEEE.

[11] Mehrnaz Ataei, Auriol Degbelo, Christian Kray, and Vitor Santos. Complying with Privacy Legislation: From Legal Text to Implementation of Privacy-Aware Location-Based Services. *ISPRS International Journal of Geo-Information*, 7(11):442, November 2018.

[12] Vanessa Ayala-Rivera and Liliana Pasquale. The Grace Period Has Ended: An Approach to Operationalize GDPR Requirements. In *2018 IEEE 26th International Requirements Engineering Conference (RE)*, pages 136–146, August 2018. ISSN: 2332-6441, 1090-705X.

[13] Kent Beck, Mike Beedle, Arie van Bennekum, Alistair Cockburn, Ward Cunningham, Martin Fowler, James Grenning, Jon Kern, Ron Jeffries, Ron Hunt, Brian Marick, Jim Highsmith, Robert C. Martin, Steve Mellor, Ken Schwaber, Jeff Sutherland, and Dave Thomas. Manifesto for Agile Software Development. 2001.

[14] Woubshet Nema Behutiye, Pilar Rodrguez, Markku Oivo, and Aye Tosun. Analyzing the concept of technical debt in the context of agile software development: A systematic literature review. *Information and Software Technology*, 82(C):139–158, 2017.

[15] France Belanger and Heng Xu. The role of information systems research in shaping the future of information privacy. *Information Systems Journal*, 25(6):573–578, 2015.

[16] S. Bellomo, N. Ernst, R. Nord, and R. Kazman. Toward Design Decisions to Enable Deployability: Empirical Study of Three Projects Reaching for the Continuous Delivery Holy Grail. In *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 702–707, June 2014.

[17] S. Bellomo, R. L. Nord, and I. Ozkaya. A study of enabling factors for rapid fielding combined practices to balance speed and stability. In *2013 35th International Conference on Software Engineering (ICSE)*, pages 982–991, May 2013.

[18] Richard Berntsson Svensson, Tony Gorschek, and Björn Regnell. Quality requirements in practice: An interview study in requirements engineering for embedded systems. In *Proceedings of the 15th International Working Conference on Requirements Engineering: Foundation for Software Quality*, REFSQ '09, pages 218–232, Berlin, Heidelberg, 2009. Springer-Verlag.

[19] Bill Kaufmann. City of Calgary apologizes for release of confidential WCB information. *Calgary Herald*, August 2016.

[20] Johann Bizer, Oliver Gnther, and Sarah Spiekermann. Technikfolgenabschtzung: Ubiquitres Computing und Informationelle. Technical report, Technische Informationsbibliothek u. Universittsbibliothek, Kiel [u.a.], 2006.

[21] Andreas Borg, Angela Yong, Pär Carlshamre, and Kristian Sandahl. The bad conscience of requirements engineering: An investigation in real-world treatment of non-functional requirements. In *Third Conference on Software Engineering Research and Practice in Sweden (SERPS'03)*, 01 2003.

[22] T. Breaux. Privacy Requirements in an Age of Increased Sharing. *IEEE Software*, 31(5):24–27, September 2014.

[23] Martin Brodin. A Framework for GDPR Compliance for Small- and Medium-Sized Enterprises. *European Journal for Security Research*, 4(2):243–264, October 2019.

[24] Frank Buschmann, David Ameller, Claudia P. Ayala, Jordi Cabot, and Xavier Franch. Architecture quality revisited. *IEEE Software*, 29(4):22–24, July 2012.

[25] Andrea Caracciolo, Mircea Filip Lungu, and Oscar Nierstrasz. How do software architects specify and validate quality requirements? In *European Conference on Software Architecture*, pages 374–389. Springer, 2014.

[26] Ann Cavoukian. Privacy by design: the definitive workshop. a foreword by ann cavoukian, ph. d. *Identity in the Information Society*, 3(2):247–251, 2010.

[27] Ilias Chantzos. GDPR Turns 1: Many Companies Still Not Ready. `https://www.symantec.com/blogs/expert-perspectives/gdpr-turns-1-many-companies-still-not-ready`, May 2019.

[28] L. Chen. Continuous Delivery: Huge Benefits, but Challenges Too. *IEEE Software*, 32(2):50–54, March 2015.

[29] Lawrence Chung and Brian A. Nixon. Dealing with non-functional requirements: Three experimental studies of a process-oriented approach. In *Proceedings of the 17th International Conference on Software Engineering*, ICSE '95, pages 25–37, New York, NY, USA, 1995. ACM.

[30] Lawrence Chung, Brian A. Nixon, Eric Yu, and John Mylopoulos. *Non-functional requirements in software engineering*. Kluwer Academic, 2000.

[31] CloudSploit. A Technical Analysis of the Capital One Hack - CloudSploit. `https://blog.cloudsploit.com/a-technical-analysis-of-the-capital-one-hack-a9b43d7c8aea`.

[32] J. Coles, S. Faily, and D. Ki-Aries. Tool-Supporting Data Protection Impact Assessments with CAIRIS. In *2018 IEEE 5th International Workshop on Evolving Security Privacy Requirements Engineering (ESPRE)*, pages 21–27, August 2018.

[33] Alison Cool. Impossible, unknowable, accountable: Dramas and dilemmas of data law. *Social Studies of Science*, 49(4):503–530, August 2019.

[34] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, March 2011.

[35] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, pages 21–21, Berkeley, CA, USA, 2004. USENIX Association.

[36] Elfriede Dustin, Jeff Rashka, and John Paul. *Automated Software Testing: Introduction, Management, and Performance*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

[37] Adam Elga. Reflection and disagreement. *Noûs*, 41(3):478–502, 2007.

[38] Neil Ernst, Stephany Bellomo, Robert L Nord, and Ipek Ozkaya. Enabling Incremental Iterative Development at Scale: Quality Attribute Refinement and Allocation in Practice. page 35.

[39] Pietro Ferrara and Fausto Spoto. Static analysis for GDPR compliance. In *Italian Conference on Cyber Security*, 2018.

[40] Brian Fitzgerald and Klaas-Jan Stol. Continuous Software Engineering and Beyond: Trends and Challenges. In *Proceedings of the 1st International Workshop on Rapid Continuous Software Engineering*, RCoSE 2014, pages 1–9, New York, NY, USA, 2014. ACM. event-place: Hyderabad, India.

[41] Brian Fitzgerald and Klaas-Jan Stol. Continuous software engineering: A roadmap and agenda. *Journal of Systems and Software*, 123:176–189, January 2017.

[42] Conner Forrest. Only 36% of firms will be fully compliant with GDPR by its deadline. `https://www.techrepublic.com/article/only-36-of-firms-will-be-fully-compliant-with-gdpr-by-its-deadline/`.

[43] Martin Fowler. Continuous integration. `https://martinfowler.com/articles/continuousIntegration.html`, 2006.

[44] Nicolas Fhnrich and Michael Kubach. *Enabling SMEs to comply with the complex new EU data protection regulation*. Gesellschaft fr Informatik, Bonn, 2019.

[45] Martin Glinz. Rethinking the notion of non-functional requirements. *Third World Congress for Software Quality*, January 2005.

[46] Martin Glinz. On non-functional requirements. pages 21–26, Oct 2007.

[47] Catarina Gralha, Daniela Damian, Anthony I. (Tony) Wasserman, Miguel Goulo, and Joo Arajo. The Evolution of Requirements Practices in Software Startups. In *Proceedings of the 40th International Conference on Software Engineering*, ICSE

'18, pages 823–833, New York, NY, USA, 2018. ACM. event-place: Gothenburg, Sweden.

[48] Casandra Grundstrom, Karin Vyrynen, Netta Iivari, and Minna Isomursu. Making Sense of the General Data Protection Regulation—Four Categories of Personal Data Access Challenges. 2019.

[49] Seda Gürses, Carmela Troncoso, and Claudia Diaz. Engineering privacy by design. *Computers, Privacy & Data Protection*, 14(3):25, 2011.

[50] Irit Hadar, Tomer Hasson, Oshrat Ayalon, Eran Toch, Michael Birnhack, Sofia Sherman, and Arod Balissa. Privacy by designers: software developers privacy mindset. *Empirical Software Engineering*, 23(1):259–289, February 2018.

[51] Johannes Heurix, Peter Zimmermann, Thomas Neubauer, and Stefan Fenz. A taxonomy for privacy enhancing technologies. *Computers & Security*, 53:1–17, 2015.

[52] Alan R Hevner, Salvatore T March, Jinsoo Park, and Sudha Ram. Design Science in Information Systems Research. page 32, March 2004.

[53] Kalle Hjerppe, Jukka Ruohonen, and Ville Leppnen. The General Data Protection Regulation: Requirements, Architectures, and Constraints. *arXiv:1907.07498 [cs]*, July 2019. arXiv: 1907.07498.

[54] Jez Humble and David Farley. *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation*. Addison-Wesley Professional, Upper Saddle River, NJ, 1 edition edition, July 2010.

[55] G. Kalogridis, M. Sooriyabandara, Z. Fan, and M. A. Mustafa. Toward unified security and privacy protection for smart meter networks. *IEEE Systems Journal*, 8(2):641–654, June 2014.

[56] Melissa Krasnow. A Summary of the California Consumer Privacy Act of 2018. `https://www.irmi.com/articles/expert-commentary/a-summary-of-ccpa-of-2018`, September 2018.

[57] Issie Lapowsky. New York's Privacy Bill Is Even Bolder Than California's. `https://www.wired.com/story/new-york-privacy-act-bolder/`, June 2019.

[58] M. Leppnen, S. Mkinen, M. Pagels, V. Eloranta, J. Itkonen, M. V. Mntyl, and T. Mnnist. The highways and country roads to continuous deployment. *IEEE Software*, 32(2):64–72, March 2015.

[59] Antonio Martini and Jan Bosch. The Danger of Architectural Technical Debt: Contagious Debt and Vicious Circles. In *2015 12th Working IEEE/IFIP Conference on Software Architecture*, pages 1–10, May 2015. ISSN: null.

[60] Steven Melendez, Steven Melendez, and Steven Melendez. A landmark Vermont law nudges over 120 data brokers out of the shadows, March 2019. Library Catalog: www.fastcompany.com.

[61] Meera Narendra. Almost a third of EU firms still not GDPR compliant. `https://gdpr.report/news/2019/07/22/almost-a-third-of-eu-firms-still-not-gdpr-compliant/`, July 2019.

[62] N. Niu, S. Brinkkemper, X. Franch, J. Partanen, and J. Savolainen. Requirements Engineering and Continuous Deployment. *IEEE Software*, 35(2):86–90, March 2018.

[63] Nicole Perlroth, Amie Tsang, and Adam Satariano. Marriott Hacking Exposes Data of Up to 500 Million Guests. *The New York Times*, November 2018.

[64] Andrew Perrin. Americans are changing their relationship with Facebook. *Pew Research Center*.

[65] Kai Petersen and Claes Wohlin. The effect of moving from a plan-driven to an incremental software development approach with agile practices. *Empirical Software Engineering*, 15:654–693, 2010.

[66] Andreas Pfitzmann and Marit Hansen. Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management A Consolidated Proposal for Terminology. page 83, 2008.

[67] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, 4(1).

[68] Eltjo R. Poort, Nick Martens, Inge van de Weerd, and Hans van Vliet. How architects see non-functional requirements: Beware of modifiability. In Björn

Regnell and Daniela Damian, editors, *Requirements Engineering: Foundation for Software Quality*, pages 37–51, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[69] Nazar Poritskiy, Flvio Oliveira, and Fernando Almeida. The benefits and challenges of general data protection regulation for the information technology sector. *Digital Policy, Regulation and Governance*, ahead-of-print, September 2019.

[70] Balasubramaniam Ramesh, Lan Cao, and Richard Baskerville. Agile requirements engineering practices and challenges: an empirical study. *Information Systems Journal*, 20(5):449–480, 2010.

[71] Edison Research. The Infinite Dial 2019. *Edison Research*, March 2019.

[72] Sandra Domenique Ringmann, Hanno Langweg, and Marcel Waldvogel. Requirements for Legally Compliant Software Based on the GDPR. In Herv Panetto, Christophe Debruyne, Henderik A. Proper, Claudio Agostino Ardagna, Dumitru Roman, and Robert Meersman, editors, *On the Move to Meaningful Internet Systems. OTM 2018 Conferences*, Lecture Notes in Computer Science, pages 258–276. Springer International Publishing, 2018.

[73] Joel Roberts. Poll: Privacy Rights Under Attack. *CBS News*.

[74] Tony Savor, Mitchell Douglas, Michael Gentili, Laurie Williams, Kent Beck, and Michael Stumm. Continuous deployment at Facebook and OANDA. In *Proceedings of the 38th International Conference on Software Engineering Companion - ICSE '16*, pages 21–30, Austin, Texas, 2016. ACM Press.

[75] Nicholas Schmidt. Top 5 Operational Impacts of CCPA: Part 5 - Penalties and enforcement mechanisms. `https://iapp.org/news/a/top-5-operational-impacts-of-cacpa-part-5-penalties-and-enforcement-mechanism`

[76] H Schulze. GDPR Compliance Report - Crowd Research Partners. `https://crowdresearchpartners.com/portfolio/gdpr-compliance-report/`, 2018.

[77] Help Net Security. Only 20% of companies have fully completed their GDPR implementations. `https://www.helpnetsecurity.com/2018/07/16/complete-gdpr-implementation/`, July 2018.

[78] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, December 2012.

[79] Devan Shah, Larry Lindsay, Josue Diaz, Sagi Shechter, and Andy Becher. Ibm security guardium analyzer bootcamp. In *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering*, CASCON '18, pages 380–382, Riverton, NJ, USA, 2018. IBM Corp.

[80] Sean Sirur, Jason R. C. Nurse, and Helena Webb. Are we there yet? Understanding the challenges faced in complying with the General Data Protection Regulation (GDPR). *arXiv:1808.07338 [cs]*, August 2018. arXiv: 1808.07338.

[81] S. Spiekermann and L. F. Cranor. Engineering privacy. *IEEE Transactions on Software Engineering*, 35(1):67–82, Jan 2009.

[82] Ossi Taipale, Jussi Kasurinen, Katja Karhu, and Kari Smolander. Trade-off between automated and manual software testing. *International Journal of System Assurance Engineering and Management*, 2(2):114–125, June 2011.

[83] Stuart A. Thompson and Charlie Warzel. Opinion | Twelve Million Phones, One Dataset, Zero Privacy. *The New York Times*, December 2019.

[84] Christina Tikkinen-Piri, Anna Rohunen, and Jouni Markkula. EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, 34(1):134–153, February 2018.

[85] Julia Carrie Wong. The Cambridge Analytica scandal changed the world but it didn't change Facebook. *The Guardian*, March 2019.