On Understanding How Developers Perceive and Interpret Privacy Requirements Research Preview

Mariana Peixoto[1(✉)], Dayse Ferreira[1], Mateus Cavalcanti[1], Carla Silva[1], Jéssyka Vilela[1], João Araújo[2], and Tony Gorschek[3]

[1] Universidade Federal de Pernambuco (UFPE), Recife, Brazil
{mmp2,dmmf,mcl2,ctlls,jffv}@cin.ufpe.br

[2] Universidade Nova de Lisboa (UNL), Lisbon, Portugal
p191@fct.unl.pt

[3] Blekinge Institute of Technology (BTH), Karlskrona, Sweden
tony.gorschek@bth.se

Abstract. [**Context and motivation**] Ensuring privacy of users' data has become a top concern in software development, either to satisfy users' needs or to comply with privacy laws. The problem may increase by the time a new law is in the vacancy period, and companies are working to understand how to comply with it. In addition, research has shown that many developers do not have sufficient knowledge about how to develop privacy-sensitive software. [**Question/problem**] Motivated by this scenario, this research investigates the personal factors affecting the developers' understanding of privacy requirements during the vacancy period of a data protection law. [**Principal ideas/results**] We conducted thirteen interviews in six different private companies. As a result, we found nine personal factors affecting how software developers perceive and interpret privacy requirements. [**Contribution**] The identification of the personal factors contributes to the elaboration of effective methods for promoting proper privacy-sensitive software development.

Keywords: Privacy requirements · Software development · Qualitative study

1 Introduction

Data handled in software applications often reveal large quantities of personal information, which are sometimes used for other purposes than initially intended and constitutes, in many cases, an invasion of privacy [6,12]. In this sense, users'

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-44429-7_8) contains supplementary material, which is available to authorized users.

N. Madhavji et al. (Eds.): REFSQ 2020, LNCS 12045, pp. 116–123, 2020.
https://doi.org/10.1007/978-3-030-44429-7_8

privacy can be defined as the right to determine when, how and to what purpose information about them is communicated to others [6].

According to Spiekermann and Cranor [10], new regulatory demands and consumer concerns are driving companies to consider privacy-friendly policies. Face to this, it is necessary to consider privacy principles and apply them from the early stages of the Software Development (SE) process, i.e., from the Requirements Engineering (RE) phase [3,6].

One approach created for this purpose is called Privacy by Design (PbD) [2]. It begins with explicit recognition of the value and benefits of proactively adopting strong privacy practices at the early stages of software development [2,5]. PbD has been embraced by the European Union to create the European General Data Protection Regulation (GDPR) [4]. This regulation was applied in May 2018 and introduced rules regarding the protection and processing of personal data. In Brazil, the General Personal Data Protection Law 13.709/2018 (in Portuguese, Lei Geral de Proteção de Dados or LGPD) was approved in August 2018 and is in the vacancy period [7].

On the other hand, there is still limited awareness of the importance of privacy requirements. For example, people are not aware of how privacy can be used to mitigate the damage caused by a potential security violation. In addition, there is little research related to the fact that developers[1] do not have sufficient knowledge of how to develop software with privacy requirements [5]. In fact, to successfully deploy PbD, we need to know how developers understand privacy [5].

In this context, we take advantage of the LGPD vacancy period, when organizations are struggling to come into compliance, to perform a qualitative study to identify the personal factors that affect how developers interpret and perceive privacy requirements in their daily work. To achieve this, we conducted thirteen semi-structured interviews with developers from six different private organizations. Data analysis was performed in light of personal factors of the Social Cognitive Theory (SCT) [1]. In SCT, a personal factor can be characterized as an element that constitutes human cognition, that is, the ability of the human being to memorize, plan, judge, among others [1,5].

Next sections are organized as follows: Sect. 2 describes the research method. Section 3 presents the study results. Section 4 details the threats to validity. And, finally, Sect. 5 shows the final considerations.

2 Research Method

We summarize the goal of our research as follows: **Analyze** personal factors, **for the purpose of** understanding their influence, **with respect to** interpretation and perception of privacy, **from the point of view of** software developers, **in the context of** Brazilian software development companies, more specifically, at Recife. Based on our goals, and a previous study provided by Hadar et al. [5],

[1] We generalize the term developer to those who work in software development.

we aim to answer the following Research Question (RQ): *What personal factors influence developers' perception and interpretation of privacy requirements in software development?*

Design and Procedures. Grounded Theory (GT) [11] was performed in light of the personal factors of SCT [1]. It is composed of the findings related to developers' perceptions and their interpretation of privacy requirements. For data collection, we performed semi-structured interviews based on the questionnaire[2] provided by Hadar et al. [5]. We decided to use the questionnaire because it was already used in previous research and validated to observe how personal factors of SCT affect the understandings of privacy by software developers. We chose non-probabilistic convenience sampling because it would be challenging to identify all members of the target population (i.e., software developers). Therefore, our candidates' selection was based on our known industrial contacts who were available and willing to participate.

We previously had a pilot interview with a member of a software development company to verify comprehension of the questions and to measure the time spent. After that, two authors conducted thirteen detailed in-depth face-to-face interviews between January 2019 and May 2019. Each interview lasted an average of 37.46 min and resulted in 8 h and 11 min of audio time. At the beginning of each interview, the participant's verbal consent, as well as audio recording permission, were confirmed to continue the procedure of data collection.

After data collection, two authors transcribed all interviews. The data analysis was conducted by four authors, based on qualitative coding principles of GT [11]. We started the coding process by performing open coding, in which we created codes for extracts of the text. After that, in axial coding, we took further readings in the transcripts and the created codes (from open coding). Thus, we identified other text extracts and also group similar codes. Finally, in selective coding, we identified categories that codes could be linked to. These categories are the personal factors that affect how developers interpret and perceive privacy in RE. We present an example of coding in Fig. 1. The coding process was performed using atlas ti software (cloud.atlasti.com).

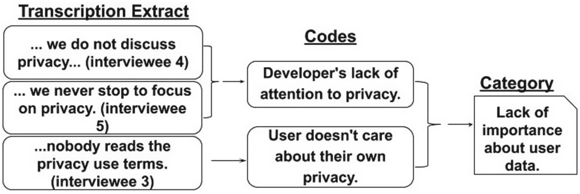
Transcription Extract
... we do not discuss privacy... (interviewee 4)
... we never stop to focus on privacy. (interviewee 5)
...nobody reads the privacy use terms. (interviewee 3)
Codes
Developer's lack of attention to privacy.
User doesn't care about their own privacy.
Category
Lack of importance about user data.

Fig. 1. Category creation.

[2] Supplementary Material: https://marianapmaia.github.io/REFSQ2020/.

3 Results and Analysis

We interviewed a total of thirteen developers from six private companies. Table 1 shows the sample characterization. The model presented in Fig. 2 explains the personal factors that play a role in developers' understanding of privacy. In the rectangles, we show nine categories as personal factors that affect positively (+) or negatively (−) how developers perceive and interpret privacy requirements. The arrows between categories (personal factors) represent that the related categories can influence each other. We also found some secondary factors (represented as a statement with an arrow to a category) which can influence positively (+), i.e., corroborate, or negatively (−), i.e., oppose the personal factors.

Table 1. Sample characterization.

Id cpy.	Cpy. size*	Domain	Role (years of experience)
1	Medium	Marketing	CEO (5)
2	Very small	Software factory	CEO (9)
3	Large	Several**	Soft. Engineer (5/5/16/10/3/4); Soft. Consultant (20)
4	Medium	Security	Soft. Analyst (3); Soft. Engineer (5)
5	Very large	Several	Developer (10)
6	Very small	Aug. reality	Developer (2)

*Number of employees: Very small < 10; Small < 100; Medium < 500; Large < 1000; Very Large > 1000. ** Offers services, maintenance, software creation, courses, etc.

Empirical knowledge about informational privacy is a positive personal factor which is corroborated by two secondary factors indicating that respondents had a practical knowledge about personal data. For example, interviewee 2 (from cpy 2) said: *"I have already served as an architect [...] that handle user data"*. This personal factor influences and is influenced by other positive personal factors. For example, **Experience in allowing the user to control their data stored by the system**, in particular, is corroborated by three secondary factors indicating that respondents concern about the need for transparency in the collection and use of personal information. For example, interviewee 12 (from cpy 3) said: *"I think all kinds of information I collect, the user has to give me consent"* .

Privacy decision depends on each development project is a positive personal factor that influences and is influenced by **Empirical knowledge about informational privacy** and **Lack of formal privacy knowledge**. This personal factor is corroborated by two secondary factors that allowed us to observe consistency among answers related to how privacy should be handled in each development project interaction. Indeed, interviewee 12 (from cpy 3) said: *"[...] it depends on each company, the way it deals with its users.*

Lack of formal privacy knowledge is a negative personal factor and it is corroborated by two secondary factors, indicating the unawareness regarding the laws and privacy definition. For example, interviewee 4 (from cpy 4) said,

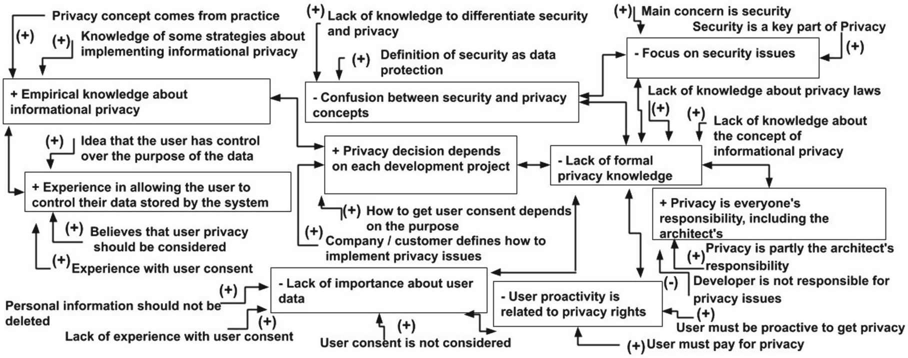
Privacy concept comes from practice
(+)
(+) Knowledge of some strategies about implementing informational privacy
+ Empirical knowledge about informational privacy
(+) Idea that the user has control over the purpose of the data
+ Experience in allowing the user to control their data stored by the system
(+) Believes that user privacy should be considered
(+) Experience with user consent
(+) Lack of knowledge to differentiate security and privacy
(+) Definition of security as data protection
- Confusion between security and privacy concepts
+ Privacy decision depends on each development project
(+) How to get user consent depends on the purpose
(+) Company / customer defines how to implement privacy issues
(+) Main concern is security
Security is a key part of Privacy (+)
- Focus on security issues
Lack of knowledge about privacy laws
(+)
(+) Lack of knowledge about the concept of informational privacy
- Lack of formal privacy knowledge
+ Privacy is everyone's responsibility, including the architect's
(+) Privacy is partly the architect's responsibility
(-) Developer is not responsible for privacy issues
- User proactivity is related to privacy rights
(+) User must be proactive to get privacy
(+) User must pay for privacy
- Lack of importance about user data
(+) Personal information should not be deleted
(+) Lack of experience with user consent
(+) User consent is not considered

Fig. 2. Personal factors influencing interpretation and perception of privacy.

"I haven't had this contact [with the law] yet". This personal factor is related to **Confusion between security and privacy concepts**, also a negative personal factor because security and privacy have different meaning. This personal factor is corroborated by two secondary factors, indicating that respondents defined privacy using security-related terms. For example, interviewee 5 (from cpy 3) said: *"I think it's the data security part, refers to the protection of personal information"*. Other answer was provided by interviewee 13 (from cpy 3): *"When you give permission to use your data, and that application eventually leaks [...] it's also a matter of privacy, but I don't know if it's a security issue"*.

Confusion between security and privacy concepts also influences and is influenced by **Focus on security issues**. This factor is corroborated by two secondary factors, indicating the respondent's main concern is just security as well as privacy is all about security. For example, interviewee 4 (from cpy 4) said: *"We need to make sure our software is secure [...]"*.

Respondents mostly believe **Privacy is everyone's responsibility, including the architect's**. One secondary factor corroborates and one opposes to this personal factor. This category showed respondents think privacy responsibility should be shared between the architect, clients, or the team. For example, interviewee 12 (from cpy 3) said: *"It is not only the responsibility of [the architect]"*. Some respondents did not believe that the responsibility for privacy lies with the developer as, for example, interviewee 12 (from cpy 3): *"Privacy issues do not come [to the developer] very much. These security issues are linked to development, but privacy issues not"*.

User proactivity is related to privacy rights is a negative personal factor with two corroborations. In some cases, it was pointed out that the right to privacy is equally proportional to the user proactivity to achieve it. Interviewee 2 (from cpy 2) quoted: *"If the application is free, you have to accept that you are the product"*. This personal factor influences and is influenced by **Lack of importance about user data**, which is also a negative factor. It has three corroborations related to the belief that data should be kept into the system regardless users' consent and privacy breach risk. For example, interviewee 12 (from cpy 3) said: *"I don't think that storing personal information is privacy violation because with this I make user's life more comfortable".*

Our findings indicate that developers have empirical knowledge of privacy, but most of them do not know how to interpret properly privacy requirements, as well as many of them do not know about formal privacy or LGPD. Empirical knowledge is a positive point, despite that, the fact of developers do not have formal knowledge can be seen as problematic because it is a period of privacy law vacancy. They generally understand that privacy could be implemented by using practices for implementing security because they make confusion between privacy and security. This finding is similar to the findings provided by Hadar et al. [5], that developers use the vocabulary of security to address privacy challenges, and this vocabulary limits their perceptions of privacy. In addition, some respondents do not intend to use privacy practices (for example, delete personal data when it is no longer needed) even recognizing their importance. They believe privacy is a trade-off, that the lack of privacy is justified by the provision of the service. Also, there was no concern to restrict the collection of personal data to only those necessary for the software operation. In fact, unrestricted data collection can become a bigger problem if a security problem occurs. This findings may be a negative factor for the acceptance and incorporation of PbD, that is, the implementation of privacy practices since the beginning of software development.

4 Threats to Validity

In the validity threats, we considered the indications provided by Runeson and Höst [9]. **Construct validity** reflects the extent to which operational measures represent what the researcher has in mind and what is investigated according to the RQs. We considered this threat by ensuring that the identities of participants and companies would not be disclosed. Besides that, prior to the interviews, we presented clarifications on the research reasons. In addition, we considered this validity when using a questionnaire already tested and validated for the same purpose (privacy point of view by developers).

Internal validity considers whether there are other factors that influence the results. To mitigate this type of threat, the sample was composed of individuals with different roles/years of experience and from companies of different sizes/domains. **External validity** is concerned with to what extent it is possible to generalize the results. We cannot assure the presented results can be generalized because the qualitative study was carried out with few participants. However, these results presented similar findings to that provided by Hadar et al. [5].

Reliability is concerned with to what extent the data and the analysis are dependent on the specific researchers. To mitigate this threat, we followed a clear method and we conducted several rounds of discussion among the involved researchers before the interviews. In addition, the interviews and data analysis were carried out by more than one researcher.

5 Final Considerations

This paper presented results of a qualitative study on how developers perceive and interpret privacy requirements. We showed nine personal factors that positively or negatively affect the developer's understanding of privacy requirements. We found that developers have practical knowledge of privacy, rather than theoretical knowledge. They often focus on security and this can compromise the resolution of privacy issues. Besides that, many developers recognize the importance of using privacy practices but some have no intention of using it.

As ongoing research, we are analysing other data collected in the interviews to observe the behavioral and environmental factors of SCT, and how they interact with personal factors and affect developers' understanding of privacy. We are also working on defining and evaluating a requirements specification method designed to guide developers to consider privacy from the beginning of agile software development [8].

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, supported by the S.E.R.T research profile, (see rethought.se, kks.se), and NOVA LINCS Research Laboratory (Ref. UID/CEC/04516/2019).

References

1. Bandura, A.: Social Foundations of Thought and Action. Prentice-Hall, Inc., Englewood Cliffs (1986)
2. Cavoukian, A.: Privacy by design: the 7 foundational principles. Inf. Priv. Commissioner Ontario Canada **5** (2009)
3. del Alamo, J.M., Martín, Y.-S., Caiza, J.C.: Towards organizing the growing knowledge on privacy engineering. In: Hansen, M., Kosta, E., Nai-Fovino, I., Fischer-Hübner, S. (eds.) Privacy and Identity 2017. IAICT, vol. 526, pp. 15–24. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92925-5_2
4. GDPR: General data protection regulation (2018). https://eugdpr.org/
5. Hadar, I., et al.: Privacy by designers: software developers' privacy mindset. Empir. Softw. Eng. **23**(1), 259–289 (2018)
6. Kalloniatis, C., Kavakli, E., Gritzalis, S.: Addressing privacy requirements in system design: the pris method. Requir. Eng. **13**(3), 241–255 (2008)
7. LGPD: General Law on Personal Data Protection/Lei Geral de Protecao de Dados n. 13.709 (2018). http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm
8. Peixoto, M., Silva, C., Lima, R., Araújo, J., Gorschek, T., Silva, J.: PCM tool: privacy requirements specification in agile software development. In: 10th Brazilian Software Conference: Theory and Practice (CBSoft 2019), pp. 108–113. SBC (2019)
9. Runeson, P., Höst, M.: Guidelines for conducting and reporting case study research in software engineering. Empir. Softw. Eng. **14**(2), 131 (2009)
10. Spiekermann, S., Cranor, L.F.: Engineering privacy. IEEE Trans. Software Eng. **35**(1), 67–82 (2008)

11. Strauss, A., Corbin, J.: Basics of Qualitative Research Techniques. Sage Publications, Thousand Oaks (1998)
12. Van Der Sype, Y.S., Maalej, W.: On lawful disclosure of personal user data: what should app developers do? In: International Workshop on Requirements Engineering and Law (RELAW), pp. 25–34. IEEE (2014)