# Privacy in Practice: Exploring Concrete Relationships Between Privacy Patterns and Privacy by Design Principles in Software Engineering

**Vinícius C. Andrade[1], Richard D. Ribeiro[1], Rafael dos P. Canteri[2], Sheila Reinehr[3], Cinthia O. de A. Freitas[3], Andreia Malucelli[3]**

[1]Federal Technology University – Paraná (UTFPR)
CEP 84017-220 – Ponta Grossa – PR – Brazil

[2]Federal University of Mato Grosso do Sul (UFMS)
CEP 79304-902 – Corumbá – MS – Brazil

[3]Pontifícia Universidade Católica do Paraná (PUCPR)
CEP 80215-901 – Curitiba – PR – Brazil

{vcandrade, richardribeiro}@utfpr.edu.br, rafael.canteri@ufms.br,
{sheila.reinehr, cinthia.freitas, andreia.malucelli}@pucpr.br

***Abstract.*** *Ensuring the fulfillment of customer preferences and requirements and adherence to legal compliance have emerged as critical considerations for software development organizations. Legislation such as the Brazillian LGPD and the European Union's GDPR highlight the importance of integrating personal data privacy rights from the beginning of system development and throughout the data lifecycle, as mentioned in the fundamental principles of Privacy by Design. However, recent studies still emphasize the need for processes, methods, guides, and tools that help translate Privacy by Design principles into practical software engineering activities. In this context, this article aims to explore the integration of abstract Privacy by Design principles into tangible Software Engineering practices. To this end, a mapping was carried out between Privacy Patterns and the principles of Privacy by Design. This process translated abstract concepts into practical activities. The reliability of the mapping process among the researchers was assessed by calculating the Intraclass Correction Coefficient (ICC). The findings underscore that when software engineers apply one or more Privacy Patterns to address personal data privacy requirements, as revealed through the correlations conducted in this study, they also tend to adhere to one or more Privacy by Design principles.*

***Keywords.*** *Privacy, Privacy by Design, Privacy Patterns, Software Engineering.*

## 1. Introduction

Concerns among data subjects regarding the collection, usage, storage, and sharing of personal information are growing [Ferrão et al. 2021]. Despite the crucial importance of

safeguarding personal data privacy, companies misusing this information persist. Notable examples include the scandal involving Facebook and Cambridge Analytica, where personal data of 87 million users was collected and misused in 2018 [Rosenberg 2018; Team 2018]. In 2021, Twitch, Amazon's streaming platform, experienced a data leak of 128 gigabytes of personal information from streamers [Browning 2021; Tidy and Molloy 2021]. In 2024, the most significant data leak in history, known as “Mother of All Breaches” (MOAB), with 26 billion user records from various services, such as Tencent, Weibo, MySpace, X (formerly Twitter), among others [Petkauskas 2024].

In the Brazilian context, an incident of data leakage occurred at Inter Bank in 2018, resulting in the loss of personal information from 19,961 account holders [Redação Veja 2018]. Subsequently, in 2021, Brazil witnessed the most extensive data leak in its history, exposing the unique identification numbers of 223 million individuals [Rohr 2021]. In 2022, 137,285 instant payment keys, known as "PIX," containing holders' data, were disclosed [Malar 2022].

Data privacy may encounter challenges across the software life cycle, potentially compromising its ultimate quality [Andrade et al. 2022; Brito et al. 2020]. Consequently, laws and regulations, including the General Data Protection Law (Lei Geral de Proteção de Dados - LGPD) [BRASIL 2018] and the European General Data Protection Regulation (GDPR) [EU 2016], emphasize the significance of incorporating Privacy by Design (PbD) principles [Cavoukian 2009]. PbD not only addresses the privacy of personal data but also clarifies the procedures for collecting and processing personal data while upholding the data life cycle from an information security perspective [Cavoukian 2009, 2012].

While implementing PbD principles in software projects is feasible, a significant level of abstraction remains [Andrade et al. 2022; Baldassarre et al. 2020; Morales-Trujillo et al. 2018; Peixoto et al. 2023]. PbD lacks the detailed guidance necessary for software engineers to seamlessly adopt regulations during the design and development of applications, thereby posing challenges to its effective implementation. As suggested by Cavoukian [2012], the next step in PbD's evolution involves "translating PbD's 7 Foundational Principles into concrete, prescriptive requirements, specifications, standards, best practices, and operational performance criteria" to provide more practical guidance for implementation.

From this scenario, which strongly impacts the entire software development cycle, the following research question arose: How can the fundamental principles of PbD be related to the best software engineering practices? To answer the research question, the PbD principles were mapped to the 72 Privacy Patterns cataloged by the University of California [UC Berkeley School of Information 2024]. In total, three privacy and personal data protection experts carried out the mapping in cycles. At the end of each cycle, they had meetings in which conflicting points were discussed to reach a consensus on the final mapping.

Integrating Privacy Patterns and Privacy by Design principles brings substantial benefits to both organizations and end users, strengthening the relationship of trust between them, as it facilitates the understanding of how abstract Privacy by Design concepts can be implemented practically in software development. This allows software engineering teams to have clear, tangible guidelines to ensure regulatory compliance, mitigate privacy risks, and meet customer expectations for personal data protection.

Furthermore, effectively enforcing Privacy Patterns increases transparency and user trust in products and services, promoting a culture of Privacy by Design.

This work is organized as follows: Section 2 presents a background on related concepts. Section 3 describes the research method. Section 4 presents the results. Section 5 discusses the relevance of the results. Section 6 presents the threats to validity, and Section 7 concludes the paper with final considerations.

## 2. Background

### 2.1. Privacy by Design

According to Cavoukian [2014], PbD is recommended for seamlessly incorporating privacy and data protection throughout the entire system development lifecycle. This integration spans from the early stages of development, extending to design, organizational processes, network architectures, and enhancements in governance systems. Additionally, PbD operates under the premise that achieving satisfactory levels of privacy protection goes beyond mere compliance with legal standards. Instead, it advocates for privacy assurance to be ingrained as a standard mode of operation [Cavoukian 2009].

PbD comprises seven foundational principles for the proactive integration of privacy considerations in the early phases of the design process [Cavoukian 2009]: (i) proactive not reactive; preventative not remedial: emphasizes a proactive stance in preventing privacy issues rather than addressing them reactively; (ii) privacy as the default: advocates for privacy protection as the automatic default, requiring explicit user choices to deviate from this setting; (iii) privacy embedded into design: calls for the seamless incorporation of privacy features directly into the design and architecture of systems; (iv) full functionality – positive-sum, not zero-sum: promotes solutions where privacy protection does not compromise system functionality, seeking positive-sum outcomes; (v) end-to-end security – lifecycle protection: advocates for security measures across the entire data lifecycle, from collection to disposal; (vi) visibility and transparency: encourages openness regarding privacy policies, practices, and procedures, ensuring transparent communication with users; and (vii) respect for user privacy: fosters a user-centric environment by empowering individuals to control their personal information and make informed decisions about its usage.

### 2.2. Privacy Patterns

According to Colesky et al. [2016], Privacy Patterns provide knowledge collected from experts in a structured, documented, and reusable way and contribute to constructing a secure information system.

The solutions offered for using these Privacy Patterns involve detailing the information assets and the level of criticality of these assets, including implementation details in a real environment, and also taking into account the architecture and technologies that must be used [Moral-García et al. 2010].

Therefore, Privacy Patterns support the documentation of standard solutions to privacy problems. They can improve systems' development by describing classes, collaborations between objects, and their purposes and help designers identify and resolve privacy-related issues [Colesky et al. 2016, 2018].

To gather proposed Privacy Patterns, researchers at the University of California maintain a catalog currently containing 72 (seventy-two) Privacy Patterns organized into categories. Each pattern is structured according to the following characteristics [Moral-García et al. 2010; UC Berkeley School of Information 2024]: (i) name: represents the problem addressed; (ii) context: contains a generic description of the configuration and specifies the conditions under which the privacy pattern should be applied; (iii) problem: presents the situation that led to the need to apply privacy mechanisms and obtain a solution; and (iv) solution: describes the solution based on the scenario and problem considered.

### 2.3. Privacy Design Strategies

Hoepman [2014] proposed a set of eight strategies that describe fundamental approaches to achieving a given objective and, consequently, satisfactory levels of privacy protection.

Hoepman's [2014] privacy strategies were constructed from existing privacy principles and personal data protection laws and are divided into two categories: Data-Driven Strategies and Driven Strategies to Processes. The first refers to privacy-friendly data processing and has the following strategies: (i) Minimise: limit the collection and processing of personal data to only what is strictly necessary for the intended purpose; (ii) Hide: protect sensitive information by implementing mechanisms to conceal or encrypt it; (iii) Separate: segregate different types of data to ensure that sensitive and non-sensitive information is stored separately; and (iv) Aggregate/Abstract: combine and analyze data in an aggregated, de-identified manner to extract insights without compromising individual privacy.

The second category highlights the processes that involve the responsible processing of personal data. It contains the following strategies: (v) Inform: keep individuals informed about the collection, use, and processing of their data through clear and transparent communication; (vi) Control: empower individuals with control over their personal information, enabling them to manage access and permissions; (vii) Enforce: implement mechanisms and policies to enforce privacy measures and ensure compliance; and (viii) Demonstrate: showcase and provide evidence of compliance with privacy principles and regulations [Hoepman 2014].

## 3. Research Method

To carry out the mapping, three privacy and personal data protection experts independently mapped the 7 (seven) fundamental principles of PbD to 72 (seventy-two) Privacy Patterns cataloged by the University of California. Figure 1 illustrates the steps of the research method.

The first stage, Planning and Preparation, has two activities: (i) Design the Mapping Protocol and (ii) Conduct Training. At this stage, each participant received a guide containing an introduction to the concept and principles of PbD [Cavoukian 2009] and a description of the introduction to Privacy Patterns containing context, problem addressed, and proposed solution [UC Berkeley School of Information 2024]. As training, 7 (seven) Privacy Patterns were chosen randomly (approximately 10% of the total). Three experts jointly mapped out the Location Granularity privacy pattern, discussing why the pattern does or does not include a specific Privacy by Design

principle. Subsequently, each researcher individually and independently mapped the remaining 6 (six) patterns: Ambient Notice, Enable/Disable Functions, Informed Consent for Web-based Transactions, Negotiation of Privacy Policy, Privacy Icons, and Protection Against Tracking.

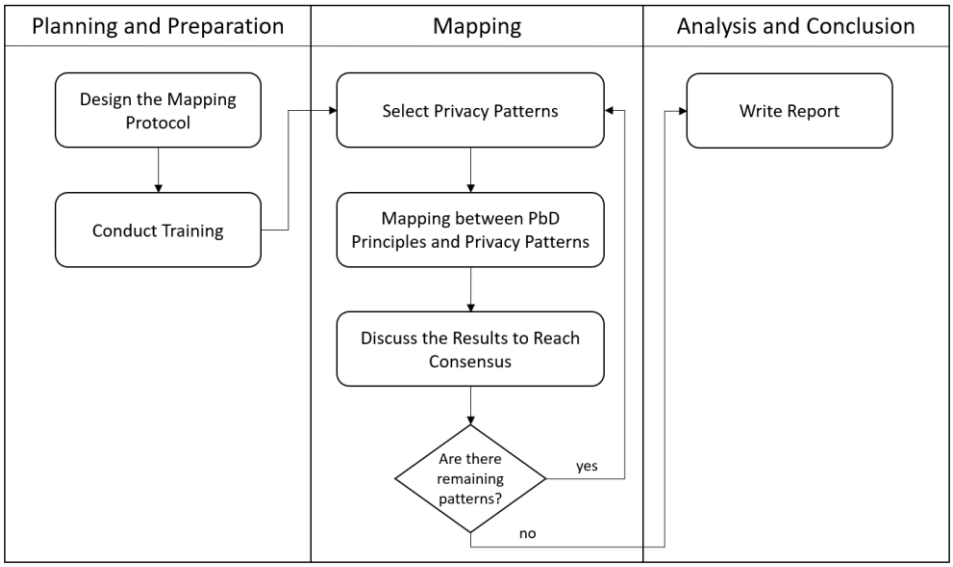


**Figure 1. Research Method**

To assess the agreement among researchers during the mapping process, we calculated the Intraclass Correlation Coefficient (ICC) [Fleiss et al. 2013]. An ICC value below 40% indicates weak agreement, while a range of 40% to 75% is considered satisfactory to good, and an ICC value of 75% or higher is deemed excellent. The ICC between researchers A and B was 77.55%, an excellent agreement rate; between researchers A and C, 79.59%, also an excellent agreement rate; and between researchers B and C, 65.31%, a satisfactory agreement rate. Following the calculation of ICC values, a meeting was convened to compare the results of each mapping. Discussions were held to reconcile any divergent findings and achieve consensus among all researchers.

The next step, Mapping, has the following activities: (i) Select Privacy Patterns, (ii) Mapping between PbD Principles and Privacy Patterns, and (iii) Discuss the Results to Reach Consensus. As this is a cyclical stage, the remaining 65 (sixty-five) Privacy Patterns were divided into 5 (five) sessions, with 13 (thirteen) patterns being chosen and mapped per session. At the end of each session, new meetings were held to discuss differences and reach a consensus among participants.

Finally, the Analysis and Conclusion stage aims to write a report containing the results obtained from the mapping carried out by the participants. At this stage, the ICC calculation was performed again. However, this time, the 72 Privacy Patterns were considered. The ICC between all researchers was considered excellent: A and B (81.55%), A and C (86.31%), and B and C (79.37%).

## 4. Results

The final mapping result between the Privacy Patterns and the Privacy by Design Principles is presented in Table 1, organized according to the corresponding Hoepman Strategies [Hoepman 2014]. The lines that display the character (●) indicate that the Privacy Pattern contemplates the Privacy by Design Principle corresponding to the column. On the other hand, its absence indicates a non-relationship.

**Table 1. Final Result of the Mapping between Privacy Patterns and PbD Principles.**

| Hoepman Design Strategies [2014] | Privacy Patterns | Privacy by Design Principles | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1. proactive not reactive | 2. privacy as the default | 3. privacy embedded into design | 4. full functionality | 5. end-to-end security | 6. visibility and transparency | 7. respect for user privacy |
| Abstract | Location Granularity | ● | ● | ● | | | | ● |
| Control | Decoupling [content] and location information visibility | ● | ● | ● | | | | ● |
| Control | Active broadcast of presence | | ● | | | | ● | ● |
| Control | Buddy List | | | | ● | | | ● |
| Control | Discouraging blanket strategies | ● | | ● | | | | ● |
| Control | Enable/Disable Functions | | | ● | | | ● | ● |
| Hide Control | Encryption with user-managed keys | ● | ● | ● | | ● | | |
| Control | Incentivized Participation | | ● | | | | | ● |
| Inform Control | Informed Consent for Web-based Transactions | | | | | | ● | ● |
| Control | Lawful Consent | ● | | ● | | | ● | ● |
| Control | Masquerade | | | ● | | ● | | ● |
| Control | Negotiation of Privacy Policy | ● | ● | | | | | ● |
| Control | Outsourcing [with consent] | ● | | | | | ● | |
| Control | Pay Back | | | ● | | | | ● |
| Control | Obtaining Explicit Consent | | | | | | ● | ● |
| Separate Control | Personal Data Store | | | ● | | ● | ● | ● |
| Control | Private link | | ● | ● | | | | |
| Control | Reasonable Level of Control | ● | | ● | | | | ● |
| Control | Reciprocity | | | ● | | | ● | ● |
| Control | Selective access control | | ● | ● | | | | ● |
| Control | Selective Disclosure | ● | ● | ● | | | | ● |
| Control | Sign an Agreement to Solve Lack of Trust on the Use of Private Data Context | | | | | | ● | ● |
| Control | Single Point of Contact | ● | | ● | | ● | | ● |
| Separate | User data confinement pattern | ● | ● | ● | | | | |
| Minimise Hide | Added-noise measurement obfuscation | ● | ● | ● | | | | |
| Hide | Aggregation Gateway | ● | ● | ● | | ● | | |
| Hide | Trustworthy Privacy Plug-in | ● | ● | ● | | ● | | |
| Hide | Anonymity Set | ● | ● | | | | | |
| Hide Separate | Anonymous Reputation-based Blacklisting | ● | ● | ● | ● | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Hide | Onion Routing | ● | ● | ● | | ● | | |
| Hide | Pseudonymous Identity | ● | ● | ● | ● | | | |
| Hide | Pseudonymous Messaging | ● | ● | ● | | ● | | |
| Hide | Use of dummies | | ● | ● | | | | |
| Hide | Attribute Based Credentials | ● | ● | | | | | ● |
| Minimise | Protection against Tracking | ● | | ● | | | | ● |
| Minimise | Strip Invisible Metadata | ● | ● | ● | ● | | | |
| Enforce | Federated Privacy Impact Assessment | ● | | ● | | | | |
| Enforce | Obligation Management | ● | | ● | | | ● | |
| Enforce | Sticky Policies | ● | | ● | | | ● | |
| Enforce | Identity Federation Do Not Track Pattern | ● | | ● | | | | |
| Inform | Abridged Terms and Conditions | | | | | | ● | |
| Inform | Appropriate Privacy Icons | | | | | | ● | |
| Inform | Ambient Notice | ● | | ● | | | ● | ● |
| Inform | Appropriate Privacy Feedback | | | | | | ● | ● |
| Inform | Asynchronous notice | | | | | | ● | ● |
| Inform | Awareness Feed | | | | | | ● | ● |
| Inform | Data Breach Notification Pattern | | | | | | ● | ● |
| Inform | Privacy Aware Wording | | | | | | ● | |
| Inform | Dynamic Privacy Policy Display | | | | | | ● | |
| Inform | Privacy icons | | | | | | ● | |
| Inform | Icons for Privacy Policies | | | | | | ● | |
| Inform | Layered Policy Design | | | | | | ● | |
| Inform | Privacy Labels | | | | | | ● | |
| Inform | Privacy Policy Display | | | | | | ● | |
| Inform | Impactful Information and Feedback | | | | | | ● | ● |
| Inform | Platform for Privacy Preferences | | | ● | | | | ● |
| Inform | Policy Matching Display | | | ● | | | | ● |
| Inform | Privacy-Aware Network Client | | | | | | ● | |
| Inform | Increasing awareness of information aggregation | | | | | | ● | ● |
| Inform | Informed Credential Selection | | | | | | ● | ● |
| Inform | Informed Secure Passwords | ● | | | | ● | | ● |
| Inform | Unusual Activities | | | ● | | | | ● |
| Inform | Informed Implicit Consent | | | | | | ● | ● |
| Inform | Minimal Information Asymmetry | ● | ● | | | | ● | ● |
| Inform | Personal Data Table | | | | | | ● | ● |
| Inform | Preventing mistakes or reducing their impact | | | | | | ● | ● |
| Inform | Privacy Awareness Panel | | | | | | ● | ● |
| Inform | Privacy Dashboard | | | | | | ● | ● |
| Inform | Privacy Color Coding | | | | | | ● | |
| Inform | Privacy Mirrors | | | | | | ● | ● |
| Inform | Trust Evaluation of Services Sides | | | | | | ● | |
| Inform | Who's Listening | | | | | | ● | ● |

The Location Granularity pattern aims to proactively prevent the holder's location data from sharing without their consent. To achieve this, the patterns objective is to minimize the collection of information by offering different levels of granularity and offering the data subject the option of how precisely they would like to share their location. Therefore, the Location Granularity pattern favors the principles of Proactive not Reactive, Privacy as the Default, Privacy Embedded into Design, and Respect for User Privacy.

Related to the Location Granularity pattern, the Decoupling [Content] and Location Information Visibility pattern allow users to decide on the sharing, disclosure, and granularity of their location-related information to use a particular service. If the user performs no configuration, the system must, by default, proactively preserve the privacy of the user's personal data. This way, the Decoupling [Content] and Location Information Visibility pattern implements the principles of Proactive not Reactive, Privacy as the Default, Privacy Embedded into Design, and Respect for User Privacy.

The Active Broadcast of Presence pattern allows the user to choose when to share their information, preventing data from being transmitted holistically in any situation, and, in case of doubts, clarifications must be provided to them. The user can choose not to be questioned again, but this decision must be made explicitly. In this context, the Active Broadcast of Presence pattern contemplates the principles of Privacy as the Default, Visibility and Transparency, and Respect for User Privacy.

Regarding social networks, users often interact with people close to them and strangers. The Buddy List pattern allows the user to maintain lists of more relevant contacts with whom they are more likely to interact. Therefore, the Buddy List pattern directly relates to the principles of Full Functionality and Respect for User Privacy.

Still, in the context of information sharing and social networks, the Discouraging Blanket Strategies pattern gives users complete control over the privacy of the shared content, enabling them to define a level of privacy that best suits their needs. As a result, the Discouraging Blanket Strategies pattern supports the principles of Proactive not Reactive, Privacy Embedded into Design and Respect for User Privacy.

For the Enable/Disable Functions pattern to be successful in its application, one must, in advance, have knowledge of which functionalities the system has and what personal information is collected in each of these functions, allowing the holder to agree or not with the collection of certain information. Therefore, the principles of Privacy Embedded in Design, Visibility and Transparency, and Respect for User Privacy are covered by the Enable/Disable Functions pattern.

To guarantee the security and privacy of holders' personal data, the Encryption With User-Managed Keys pattern aims to encrypt holders' personal information before storing it or transferring it through online services. The principles Proactive not Reactive, Privacy as the Default, Privacy Embedded into Design, and End-to-End Security are considered in the Encryption With User-Managed Keys pattern.

Users' participation in a given system by providing personal data to improve the identification of their preferences is necessary for data controllers. In this context, the Incentivized Participation pattern encourages user participation without harming or invading the privacy of data subjects. Any activity that requires collecting user data

requires prior consent. Therefore, the Incentivized Participation pattern is related to Privacy as the Default and Respect for User Privacy principles.

To maintain a profitable service and provide a better user experience, controllers need to collect personal data from data subjects. However, collection is only permitted with the data subject's informed consent. Therefore, the Informed Consent for Web-based Transactions privacy pattern aims to protect the interests of data subjects and establishes that the collection of personal information can only occur with their consent, and this can only happen after the presentation of clear and concise information about how data will be collected, stored, processed and deleted. Therefore, the Informed Consent for Web-based Transactions pattern favors the principles of Visibility and Transparency and Respect for User Privacy.

In this same context of obtaining consent from data subjects, the Lawful Consent pattern establishes that services must be separate and, for each specific service, explicit consent must be acquired, preventing the user from sharing their personal data without prior knowledge of a particular functionality. In this way, the Lawful Consent pattern covers the principles of Proactive not Reactive, Privacy Embedded into Design, Visibility and Transparency, and Respect for User Privacy.

The Mascarade pattern aims to allow the user to select the level of sharing of personal information for a given context. It is accomplished by organizing personal information into privacy scales. When the user selects a level on the scale, all information from that and lower levels is shared. In this context, the Mascarade pattern includes Privacy Embedded into Design, End-to-End Security, and Respect for User Privacy principles.

The Negotiation of Privacy Policy pattern proposes that users' privacy preferences, when not recognized, are permanently configured to preserve their privacy as much as possible. This care must be taken by incorporating techniques that protect the user's privacy from the beginning of using the service. It is done by implementing a data leakage restriction until the user allows which information can be shared, thus respecting their privacy. In this context, the Negotiation of Privacy Policy pattern involves the principles of Proactive not Reactive, Privacy as the Default, and Respect for User Privacy.

In some cases, controllers need to share data with third parties to process the holder's data. The Outsourcing [with consent] pattern establishes that, in these cases, controllers must transparently present to data subjects what the data is and how third parties will process it. Third-party data processing will only be permitted after obtaining the data subject's free, specific, and explicit consent. This way, the Outsourcing [with consent] pattern implements the Proactive not Reactive, and Visibility and Transparency.

The Pay Back pattern benefits users when they contribute or maintain content for the service, such as on a social network. However, the individual choices of users must be respected, and for those who choose to provide more information, the legal consent of the holder must be obtained. Therefore, the Pay Back pattern is related to the principles of Privacy Embedded in Design and Respect for User Privacy.

The Obtaining Explicit Consent pattern aims to provide the data subject with clear and objective notifications about how a particular service will collect, process, and

store its data. The controller must ensure the holder understands the information and consequences of accepting the presented terms. Consent must be given freely by the holder. Therefore, the Obtaining Explicit Consent pattern supports the principles of Visibility and Transparency, and Respect for User Privacy.

The Personal Data Store pattern discusses combining a central server and personal tokens. In this pattern, tokens can take on USB keys and incorporate a database system and an authentication certificate for the holder. As a result, holders have greater control over their personal data, which remains secure and stored locally and can be maintained by the holder. Therefore, the Personal Data Store pattern contemplates the principles of Privacy Embedded into Design, End-to-End Security, Visibility and Transparency, and Respect for User Privacy.

The Private Link pattern aims to provide the holder with a private link to a specific resource, such as personal information about the user. If deemed necessary, the holder may share the link with others, giving them access to that personal information. Therefore, the Private Link pattern favors the principles of Privacy as the Default and Privacy Embedded into Design.

To allow users to provide information in a selective and granular manner, the Reasonable Level of Control pattern aims, in advance, to design ways to enable the holder to choose their data's privacy level, granting third parties access to the data. In this context, the Reasonable Level of Control pattern is related to the Proactive not Reactive, Privacy Embedded into Design, and Respect for User Privacy.

In systems where users must collaborate among themselves in favor of the group, the quality of the final result is given through the contribution of all members. The Reciprocity pattern mentions that each member must be rewarded in proportion to their participation in the group's earnings. Any collection and use of user data must be consented to by the user and informed of how this personal data will be used, respecting their privacy. Therefore, the Reciprocity pattern includes the principles of Privacy Embedded into Design, Visibility and Transparency, and Respect for User Privacy.

The Selective Access Control pattern gives users control over the visibility of content shared in social environments. Users can specify rules based on other users or groups of people to define who to target the post. Therefore, the Selective Access Control pattern implements the principles of Privacy as the Default, Privacy Embedded into Design, and Respect for User Privacy.

Occasionally, users wish to use a service anonymously, thus minimizing the chances of providing their personal data. On the other hand, controllers need some information, as this way, users cannot carry out malicious activities without being identified. For these cases, the Selective Disclosure pattern identifies which information is essential for the system to function, thus promoting data minimization. Furthermore, anonymous functionality must be provided to the user as long as it does not compromise the service. Therefore, the Selective Disclosure pattern considers the principles of Proactive not Reactive, Privacy as the Default, Privacy Embedded into Design, and Respect for User Privacy.

For users' personal data to be processed, the controller needs the consent of the data subjects. In this context, the Sign an Agreement to Solve Lack of Trust on the Use of Private Data Context pattern establishes that the service must present to the user

mechanisms through which data will be collected and how it will be processed, in addition to binding to the controller or its representatives. This information must be presented clearly and objectively, and data collection can only be carried out after obtaining the user's consent. In this way, the Sign an Agreement to Solve Lack of Trust on the Use of Private Data Context pattern encompasses the principles of Visibility and Transparency, and Respect for User Privacy.

Distributed storage services require specialized privacy management. The Single Point of Contact pattern adopts an approach by providing security tokens to authenticate and authorize the presentation of confidential information to a given user and storing pseudonyms to preserve the data subject's identity. Therefore, the Single Point of Contact pattern supports the principles of Proactive not Reactive, Privacy Embedded into Design, End-to-End Security, and Respect for User Privacy.

When the collection of personal data represents a threat to the privacy of data subjects, the system architecture can be rethought. The User Data Confinement Pattern establishes a change in the trust relationship between data subjects and service providers. Instead of the holders' personal data being collected and processed by the service provider, the processing occurs locally in the user's trusted environment. It allows for greater control of personal data by the holders. In this context, the User Data Confinement Pattern contemplates the Proactive not Reactive, Privacy as the Default, and Privacy Embedded into Design principles.

The Added-Noise Measurement Obfuscation pattern aims to add a noise value to the actual value of an attribute collected from the data subject before transmitting it. Thus, when obtaining the data, a third party would not be able to infer the correct value of it, preserving the data subject's privacy. In this way, the Added-Noise Measurement Obfuscation pattern is related to the Proactive not Reactive, Privacy as the Default, and Privacy Embedded into Design principles.

## 5. Discussion

The mapping of Privacy by Design principles with Privacy Patterns resulted in a varied distribution of relationships, highlighting nuances in applying these principles in software engineering. Analysis of relationship numbers reveals significant differences between principles, raising intriguing questions about the underlying reasons for these disparities.

Respect for User Privacy, with 42 relationships, was the principle with the most relationships. Linking this principle to many patterns highlights its pervasive relevance and the centrality of respect for user privacy in software development practices. Furthermore, this principle is related mainly to Privacy Patterns categorized as Control and Inform, which suggests that promoting user autonomy and transparency are central approaches to manifesting this principle in software engineering.

The second Privacy by Design principle with the most relationships was Visibility and Transparency, related to 39 Privacy Patterns. The high number of patterns associated with this principle suggests that the emphasis on providing visibility and transparency has broad applications and is considered essential in various privacy circumstances. We note that the principle of Visibility and Transparency is predominantly associated with Privacy Patterns identified as Inform. It reflects the

importance of transparent communication to ensure users' understanding, trust, and active participation in privacy practices.

Next, with 34 relationships, the Privacy Embedded into the Design principle is due to several Privacy Patterns focus on the proactive integration of privacy in the early stages of design and reflect a commitment to controlling and ensuring compliance with privacy practices throughout the software lifecycle. For this reason, this principle is mainly related to Privacy Patterns grouped into the Control, Hide, and Enforce categories.

The Proactive not Reactive principle was related to 29 Privacy Patterns as it emphasizes, from the beginning of development, anticipating threats, minimizing risks, hiding sensitive data, preventing privacy violations, and ensuring legal compliance with privacy issues. It explains the tendency to associate the principle with privacy patterns grouped under the Minimise, Hide, and Enforce categories.

Subsequently, the Privacy as the Default principle was associated with 22 Privacy Patterns. Linking this principle to fewer patterns suggests that although privacy is considered from the outset, the range of practical implementations is restricted. The limited collection of personal data combined with the concealment of information and a cautious and preventive approach in managing the data subject's privacy means that the Privacy as the Default principle has a relationship with privacy patterns mainly classified as Hide and Minimise.

Finally, the principles with the lowest number of associations are End-to-End Security (9 associations) and Full Functionality (4 associations). Despite being a fundamental pillar, End-to-End Security is not directly associated with various privacy scenarios. In turn, implementing the Full Functionality is challenging due to the complexity inherent in balancing full functionality delivery with the need to protect users' privacy in compliance with various regulations and individual expectations.

With the mapping between the Privacy Patterns and the Privacy by Design Principles, there is a complexity in correlating them, as, sometimes, the abstract nature of the Privacy by Design principles generates doubts when relating them to the Privacy Patterns. Subsequently, these doubts were resolved in meetings with the other participants, in which there was an explanation to reach a consensus on the correlation between certain Privacy Patterns and a specific Privacy by Design principle.

Furthermore, we observed a conceptual similarity between Privacy Patterns and Privacy by Design principles, thus linking practice with theory. As a result of the mapping, we obtained an instrument that can assist software engineers in making decisions and implementing specific privacy problems considering the Privacy by Design principles required by legislation and regulations.

## 6. Threats to Validity

In this work, a threat to the validity of the construct lies in the interpretation and definition of Privacy Patterns. Differences in understanding among researchers in the relationship between Privacy by Design principles and Privacy Patterns may impact the consistency of the mapping. Another threat relates to the interpretation of the Privacy by Design principles, as they can be interpreted differently. Divergences in researchers' understanding of these principles may influence the attribution of Privacy Patterns to

specific principles. Discussions between researchers were held before and throughout the mapping to minimize this threat, aiming at reaching a consensus.

The threat to external validity can be cited as the generalization of the results. The research focused on Privacy Patterns cataloged by the University of California, which may limit the generalization of results to other sets of patterns used in different contexts or countries. This threat can be considered minimal once the used catalog is quite comprehensive.

Finally, reliability between researchers can be mentioned as a threat to internal validity. This threat was mitigated by calculating the Intraclass Correlation Coefficient (ICC), demonstrating satisfactory and excellent agreement between researchers.

## 7. Conclusions

The present study sought to investigate how the fundamental principles of Privacy by Design [Cavoukian 2009] can be related to better software engineering practices, recognizing the challenges highlighted by several authors regarding the difficulty of translating these highly abstract principles into practical software development activities [Andrade et al. 2022; Baldassarre et al. 2020; Morales-Trujillo et al. 2018; Peixoto et al. 2023]. Faced with this complexity, our approach aimed to create a bridge between principles and tangible actions for developers using the 72 Privacy Patterns cataloged by the University of California [UC Berkeley School of Information 2024].

This initiative aims to bridge the gap between theory and practice and provide software development teams with a solid and applicable framework for integrating privacy considerations from the earliest stages of the software development process, such as in writing user stories by identifying the personal data that will be collected or requesting the holder's consent to carry out a specific action. At this point, it is possible to use mapping to define which privacy patterns will be used and, consequently, which PbD principles will be related.

The detailed mapping of these Privacy Patterns was conducted by three privacy and personal data protection experts, who established relationships between each pattern and one or more Privacy by Design principles. We calculated the Intraclass Correction Coefficient (ICC) [Fleiss et al. 2013] to ensure reliability and consistency in the results. The agreement rates between all the researchers were considered excellent [Fleiss et al. 2013].

When faced with issues related to personal data privacy, software engineers can use one or more Privacy Patterns related to one or more Privacy by Design principles to ensure a systematic and practical approach to resolving these issues. By applying Privacy Patterns in accordance with Privacy by Design principles, software engineers ensure adherence to privacy regulations and promote user trust, transparency, and accountability in handling personal data. This proactive approach from the early phases of software development helps avoid compliance issues and potential legal sanctions. It contributes to building products and services that respect privacy and meet users' expectations. In this way, the mapping between Privacy Patterns and Privacy by Design principles is instrumental in guiding software engineering practices toward a safer, more ethical, and trustworthy digital environment, benefiting both organizations and individuals.

As future work, we are currently developing a software process focused on protecting the privacy of personal data. This process is designed to systematically integrate Privacy by Design principles from the early stages of development. Our goal is to offer clear and practical guidelines for software engineers, enabling them to seamlessly integrate privacy considerations at every stage of the software lifecycle. We are additionally in the process of creating a tool designed to aid software engineers in accessing pertinent information regarding Privacy by Design and Privacy Patterns. This tool will consist of an interactive platform equipped with educational materials, real-world examples, and customized guidelines tailored to selected patterns. By simplifying the implementation process and fostering a culture of privacy-conscious development, this tool aims to facilitate the integration of privacy considerations into software engineering practices.

## References


Andrade, V. C., Gomes, R. D., Reinehr, S., Freitas, C. O. D. A. and Malucelli, A. (2022). Privacy by Design and Software Engineering: a Systematic Literature Review. In *Proceedings of the XXI Brazilian Symposium on Software Quality*.

Baldassarre, M. T., Santa Barletta, V., Caivano, D. and Scalera, M. (2020). Integrating security and privacy in software development. *Software Quality Journal*, p. 1–32.

BRASIL (2018). Lei Geral de Proteção de Dados Pessoais (LGPD). http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm, [accessed on Oct 9].

Brito, I. S., Moreira, A. and Araújo, J. (2020). Handling Nonfunctional Requirements for Smart Cities. In *23rd Iberoamerican Conference on Software Engineering, CIbSE 2020*.

Browning, K. (2021). A “potentially disastrous” data breach hits Twitch, the livestreaming site. https://www.nytimes.com/2021/10/06/technology/twitch-data-breach.html, [accessed on Jun 7].

Cavoukian, A. (2009). Privacy by Design - The 7 foundational principles - Implementation and mapping of fair information practices. *Information and Privacy Commissioner of Ontario, Canada*, p. 5.

Cavoukian, A. (2012). Operationalizing Privacy by Design: A Guide to Implementing.

Cavoukian, A., Shapiro, S. and Cronk, R. J. (2014). *Privacy engineering: Proactively embedding privacy, by design*. Office of the Information and Privacy Commissioner.

Colesky, M., Caiza, J. C., Del Lamo, J. M., Hoepman, J. H. and Martín, Y. S. (2018). A system of privacy patterns for user control. *Proceedings of the ACM Symposium on Applied Computing*, p. 1150–1156.

Colesky, M., Hoepman, J. and Hillen, C. (2016). A Critical Analysis of Privacy Design Strategies. In *2016 IEEE Security and Privacy Workshops (SPW)*.

EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. *Official Journal of the European Union (OJ)*, v. 59, n. 1–88, p. 294.

Ferrão, S. É. R., Carvalho, A. P., Canedo, E. D., et al. (2021). Diagnostic of Data Processing by Brazilian Organizations - A Low Compliance Issue. *Information*, v. 12, n. 4, p. 1–30.

Fleiss, J. L., Levin, B. and Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.

Hoepman, J. H. J.-H. (2014). Privacy design strategies. *IFIP Advances in Information and Communication Technology*, v. 428, p. 446–459.

Malar, J. P. (2022). Banco Central anuncia vazamento de dados ligados a mais de 130 mil chaves Pix. https://www.cnnbrasil.com.br/economia/banco-central-anuncia-vazamento-de-dados-ligados-a-mais-de-130-mil-chaves-pix/, [accessed on Jan 25].

Moral-García, S., Ortiz, R., Moral-Rubio, S., Vela, B., Garzás, J. and Fernández-Medina, E. (2010). A new pattern template to support the design of security architectures. In *The Second International Conferences of Pervasive Patterns and Applications*.

Morales-trujillo, M. E., Matla-cruz, E. O., García-Mireles, G. A. and Piattini, M. (2018). Privacy by Design in Software Engineering: a Systematic Mapping Study. In *Ibero-American Conference on Software Engineering (CIbSE)*.

Peixoto, M., Ferreira, D., Cavalcanti, M., et al. (2023). The Perspective of Brazilian Software Developers on Data Privacy. *Journal of Systems and Software*, v. 195, p. 111523.

Petkauskas, V. (2024). Mother of all breaches reveals 26 billion records: what we know so far. https://cybernews.com/security/billions-passwords-credentials-leaked-mother-of-all-breaches/#:~:text=Mother of all breaches reveals,what we know so far&text=Image by Cybernews.,mind-boggling 26 billion records., [accessed on Feb 6].

Redação Veja (2018). Banco Inter vai pagar R$ 1,5 milhão por vazamento de dados de clientes. https://veja.abril.com.br/economia/banco-inter-vai-pagar-r-15-milhao-por-vazamento-de-dados-de-clientes/, [accessed on May 24].

Rohr, A. (2021). Megavazamentos de dados expõem informações de 223 milhões de números de CPF. https://g1.globo.com/economia/tecnologia/blog/altieres-rohr/post/2021/01/25/vazamentos-de-dados-expoem-informacoes-de-223-milhoes-de-numeros-de-cpf.ghtml, [accessed on May 24].

Rosenberg, M. (2018). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html, [accessed on Mar 10].

Team, C. 4 N. I. (2018). Revealed: Cambridge Analytica data on thousands of Facebook users still not deleted. https://www.channel4.com/news/revealed-cambridge-analytica-data-on-thousands-of-facebook-users-still-not-deleted, [accessed on Mar 10].

Tidy, J. and Molloy, D. (2021). Twitch confirms massive data breach. https://www.bbc.com/news/technology-58817658, [accessed on Jun 7].

UC Berkeley School of Information (2024). Privacy Patterns. https://privacypatterns.org/, [accessed on Feb 6].

Conceitos chave: Privacy By Design, Privacy Patterns, Hopeman's strategies.