Privacy by Design in Software Engineering: An update of a Systematic Mapping Study

The 38th ACM/SIGAPP Symposium On Applied Computing - SAC 2023

Shirlei Aparecida de Chaves, Fabiane Barreto Vavassori Benitti

Postgraduate Program in Computer Science (PPGCC) - Federal University of Santa Catarina (UFSC)

March 27 - March 31, 2023

Motivation

Why is PbD in SE of interest?

Figure: Several Regulations went into effect from 2018 on

- Regulations aim to protect personal data
- Privacy by Design (PbD) is core concept - has criticism for the lack of methodology for SE practice

This study aims to understand SE's focus on PbD challenges.

Motivation

Why an update?

- A previous Systematic Mapping Study (SMS)[1] conducted in 2017-2018 found limited support for embedding privacy into software development, with most research focus on privacy patterns and requirements.
  - before GDPR enforcement

[1] Miguel Ehecatl Morales-Trujillo, Gabriel Alberto García-Mireles, Erick Orlando Matla-Cruz, and Mario Piattini. 2019. A Systematic Mapping Study on Privacy by Design in Software Engineering. CLEI Electronic Journal 22, 1 (April 2019). https://doi.org/10.19153/cleiej.22.1.4

Motivation

Why an update?

- The current study was designed to update the original SMS in order to verify the extent to which the software engineering community has adapted to address the challenges of Privacy by Design since the enforcement of GDPR and the emergence of new technologies.
  - topic is still relevant to policy and practice
  - preliminary searches suggested that there are new studies suitable for inclusion

Main Contributions

- Visualization of the changes in the state of the art regarding PbD in SE, especially after GDPR full enforcement
- Overview of research gaps and trends of PbD in SE, summarised as research directions.

What is the State of the Art of PbD when applied to software engineering?

What is the State-of-the-Art of PbD when applied to software engineering?

- RQ1.What is the meaning of PbD in the context of SE?
- RQ2.What privacy goals have been addressed in the development of methodological support for SE?
- RQ3.What privacy principles were addressed in the selected papers?
- RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

What is the State-of-the-Art of PbD when applied to software engineering?

- RQ1.What is the meaning of PbD in the context of SE?
- RQ2.What privacy goals have been addressed in the development of methodological support for SE?
- RQ3.What privacy principles were addressed in the selected papers?
- RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

What is the State-of-the-Art of PbD when applied to software engineering?

- RQ1.What is the meaning of PbD in the context of SE?
- RQ2.What privacy goals have been addressed in the development of methodological support for SE?
- RQ3.What privacy principles were addressed in the selected papers?
- RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

What is the State-of-the-Art of PbD when applied to software engineering?

- RQ1.What is the meaning of PbD in the context of SE?
- RQ2.What privacy goals have been addressed in the development of methodological support for SE?
- RQ3.What privacy principles were addressed in the selected papers?
- RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

What is the State-of-the-Art of PbD when applied to software engineering?

- RQ1.What is the meaning of PbD in the context of SE?
- RQ2.What privacy goals have been addressed in the development of methodological support for SE?
- RQ3.What privacy principles were addressed in the selected papers?
- RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

Methodology

Search and selection strategy

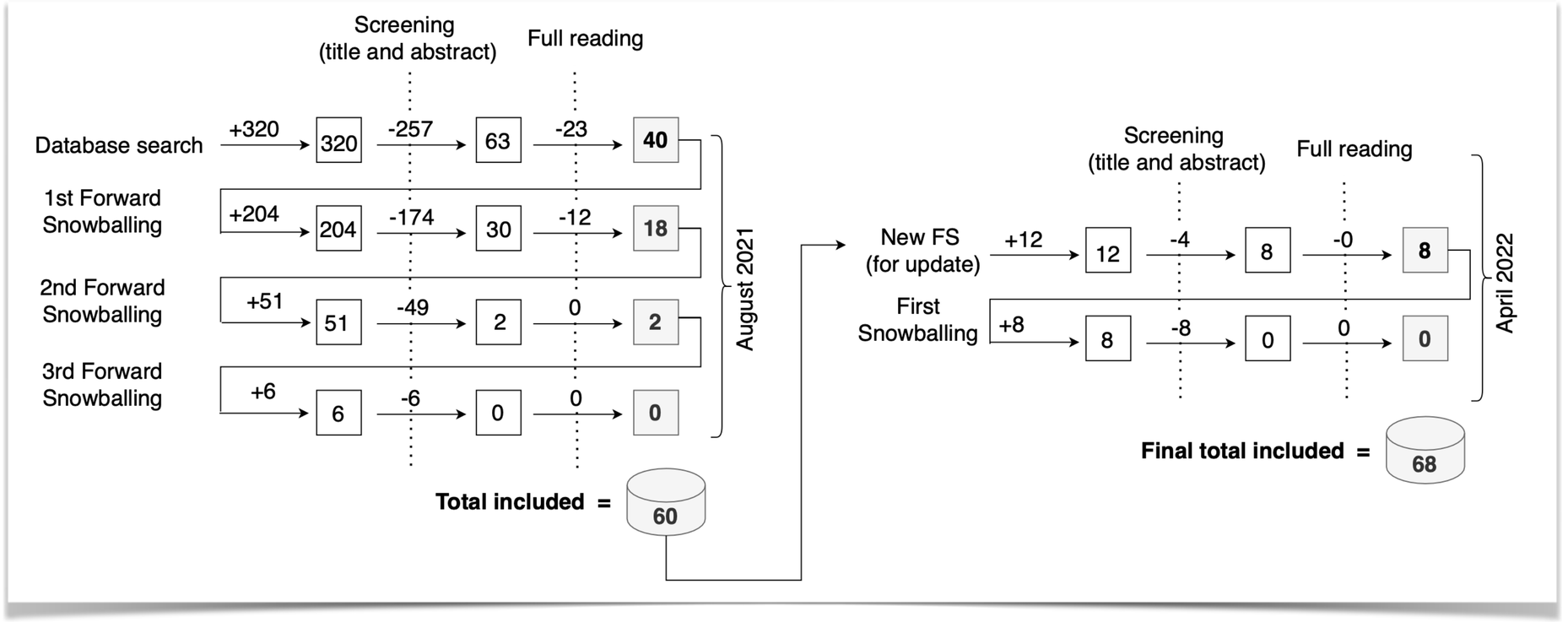
Databases: IEEE Xplore, ACM, Scopus

Search string:
("privacy by design") AND ("software engineering" OR "information systems" OR "requirements engineering")

- **Inclusion criteria**: The paper discusses or apply PbD in SE, published after January 2018, peer-reviewed, written in English, full text available.
- **Exclusion criteria**: Duplicated papers, not in English, editorials, tutorials, PhD dissertations, or Master theses.

Methodology

Search and selection strategy



Results

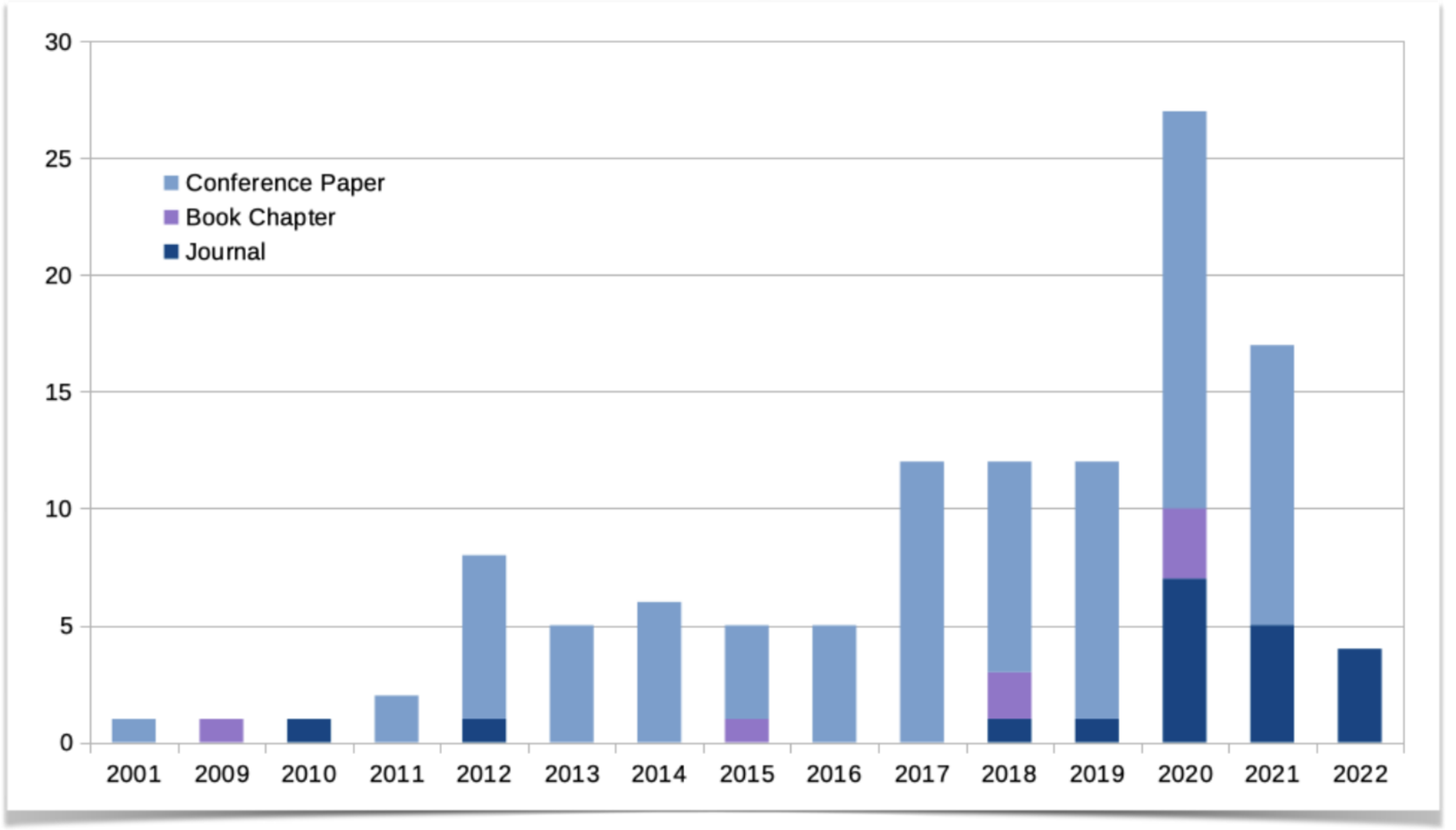
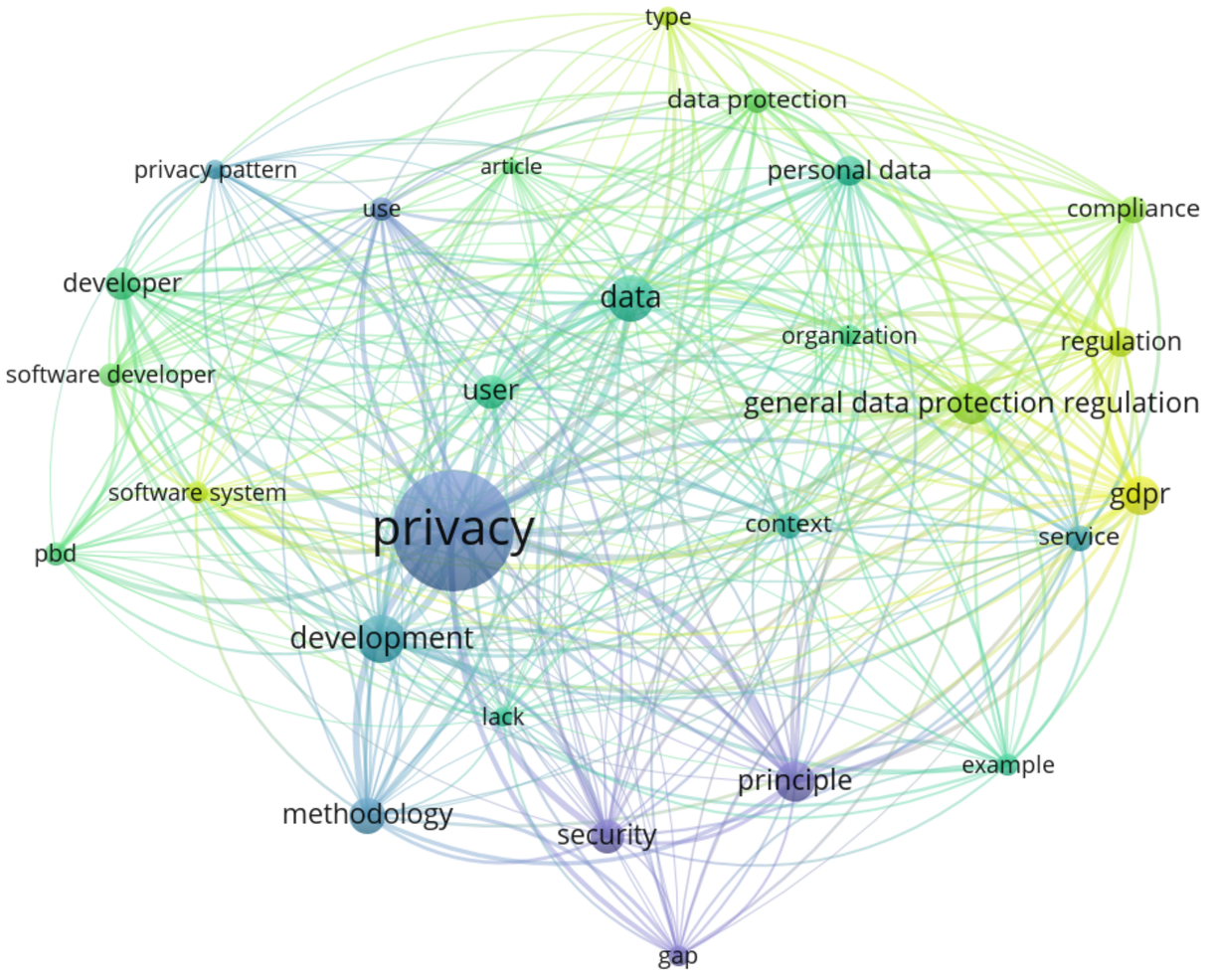


Figure: Paper contribution type by year

Results



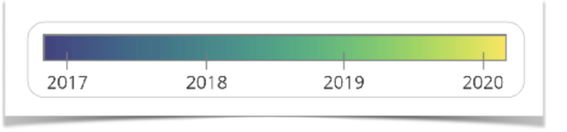


Figure: Overlay visualization based on terms weight. The gradient color from blue to yellow indicates average per year.

Results

RQ1. What is the meaning of PbD in the context of SE?

- Original SMS: 57% of the papers presented a definition of PbD, by establishing its definition or determining its goals;
- Update: 61% did not try to define PbD, but considered GDPR and Cavoukian's PbD

It seems that PbD meaning has reached a consensus around GDPR definitions.

Results

RQ2.What privacy goals have been addressed in the development of methodological support for SE?

- Original SMS: Data Minimisation (DM) was the most recurring goal.
- Update: no privacy goal standing out

Compliance with privacy regulations in general.

Results

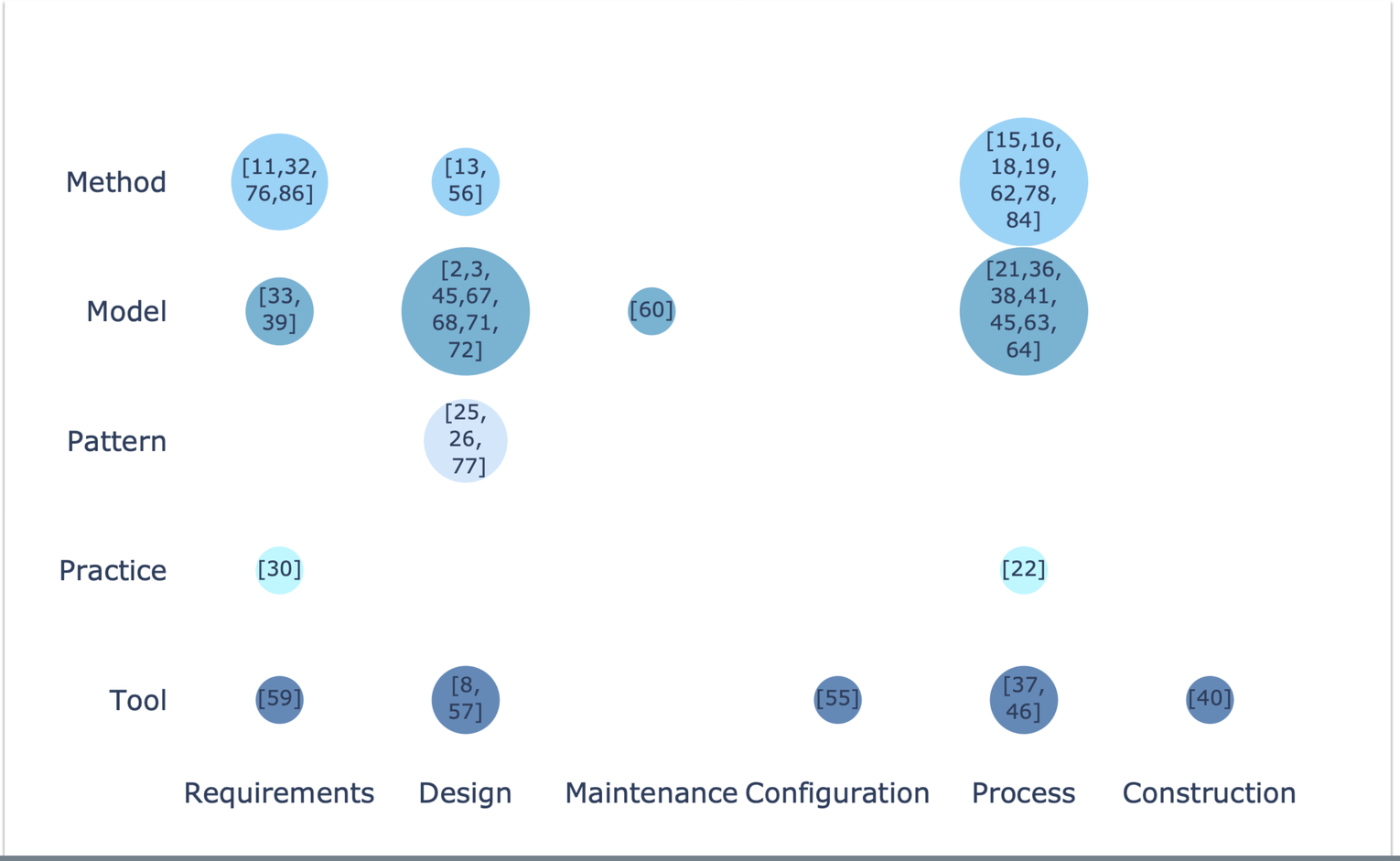
RQ3.What privacy principles were addressed in the selected papers?

- Original SMS: GDPR (36%), OECD (10%), ISO/IEC 29100 (18%), FIPPS (13%) and others (15%)
- Update: GDPR (82%), one paper cites ISO 29110 and the remainder no particular source of principles.

The emphasis on GDPR is expected because it is a regulation that leverages privacy principles from sources like the FIPPs and has become a source of concern.

Results

RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?



Results - Models

RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

- Integrating privacy by design (PbD) into the software development life cycle (SDLC) with the W-Model, Privacy-Aware SDLC, and a framework integrated into SDLC.
- Measuring privacy risk and enhancing system design based on privacy controls, strategies, and cost models.
- Facilitating software development and adaptation with a conceptual information flow model and a meta-model to manage knowledge.
- Ensuring privacy through architectural approaches in service-oriented architecture
- Bridging the gap between legal text and technical context with data sovereignty governance framework, variability aware legal-GRL, model-based privacy analysis, and an architectural viewpoint for data protection

Results - Methods

RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

- Agile Methods: Human Centered Design Security Scrum (HCD-Security Scrum) and Threat Poker game for estimating security and privacy risks.
- Complementing existing methods: Privacy Oriented Software Development (POSD) and Privacy by Evidence (PbE) methodology.
- Requirement Engineering (RE): lightweight method for elicitation of privacy and data protection requirements, P-STORE for privacy objectives, and extending BPMN for GDPR compliance evaluation.
- Privacy threat modeling: early lightweight privacy analysis approach and formalizing norms as executable specifications.
- End-user perspective: A/P Test for information systems privacy.

Results - Patterns

RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

- Enterprise architecture patterns for GDPR compliance
- Selection-support method for privacy patterns
- Empirical study of privacy patterns in design process

Results - Tools

RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

- Annotating Abstract Syntax Tree with personal data
- Analyzing compliance: web applications, GDPR
- Transparency in RESTful applications, with agile and DevOps practices
- Enhancing Requirements Engineering tools using gamification
- Designing privacy-aware IoT applications
- Integrating privacy in agile development methods
- Using Privacy Knowledge Base for decision support
- Privacy-Aware Data Flow Diagrams

Results - Practice

RQ4.What approaches for enhancing privacy in the context of SE have been proposed in the selected papers?

- Developers’ intention to use Privacy Enhancing Methodologies (PEM)
- Obstacles faced in GDPR implementation and developing privacy-preserving software
- Factors afecting developer’s understanding and adressing of privacy
- Importance of understanding developers’ attitudes towards data monetisation
- Development of a scale to measure developers’ attitudes towards handling personal data

Validity Threats

- Descriptive validity
- Theoretical validity
- Generalizability
- Interpretive validity
- SMS update-specific threats

Research Directions

- User-centric approach
- Focus on Construction, Maintenance, Configuration and Testing
- Industrial settings

Conclusions

Original SMS: little support for embedded privacy during software development

Update: Recent changes observed with increase in proposals for models and methods to address activities in multiple stages. Focus on software developers and their challenges in promoting PbD.

Research gaps remain, including user-centric approaches and validation in industrial settings

Final remarks

List of selected papers and Data Extraction Form (DEF) are available at:
https://github.com/shirlei/pbd-se-sms-update

Email: shirlei@gmail.com

Thank you!