Sok: A Systematic Review on the Evaluation of Privacy Risk Assessments and Privacy Impact Assessment Methodologies - Review Protocol

Research team: Anonymous

Document Version Control			
Document status	**Version**	**Changes from previous version**	**Date**
Initial draft	v0.1	- Proposed protocol for SLR - First feedback based on discussion incorporated	2022-11-22
Second draft	v0.2	- Search strings improved - Title improved to the one in the current version - Protocol re-written in detail.	2022-12-21
Final draft	v1.0	- Search strings improved and agreed - Research questions specified in detail - Review protocol agreed - Refined the exclusion criteria (2023-02-07)	2023-01-05

Objective:

Privacy Impact Assessments (PIAs), also known as Data Protection Impact Assessments (DPIAs), are systematic processes used for identifying privacy threats and selecting appropriate safeguards when designing working systems that process personal data. As part of this process, a specific method can be used for the Privacy Risk Analysis (PRA), which is often considered as a core component of a PIA methodology.

Although many PIA methodologies have been proposed in the last years, most of them come without strong scientific evidence of their validity and reliability other than in terms of a limited evaluation and comparative analysis. For example, the works of (Wadhwa and Rodrigues, 2013; Wright, Finn and Rodrigues, 2013; Vemou and Karyda, 2018) compare approaches in terms of associated legal frameworks, scope and depth of existing PIA methodologies.

Given that, this project aims to conduct a Systematic Literature Review (SLR) to identify the available PIA and PRA methodologies in published research and to appraise their maturity in terms of scientifically rigorous evaluation strategies.

Rationale

From a preliminary search, it was noted that there is lack of systematic reviews on the topic of PIAs and PRAs. We conducted a prior search to rule out the possibility of duplicating existing reviews or finding reviews that relate to this SLR's topic by using the search string below:

("review" AND "privacy impact assess*" OR "data protection impact assess*" OR "data protection risk assess*" OR "privacy risk analys*")

The need to determine whether PIA methodologies (including PRA methodologies) have been evaluated is necessary, despite the fact that the field of PIA has received a lot of attention (e.g., on PIA evaluation criteria and comparative analysis of PIA tools (Wadhwa and Rodrigues, 2013), evaluation of PIA guidance documents (Clarke, 2011), etc.).

For such reasons, a systematic review on the topic is warranted, especially to identify existing solutions for PIAs and PRAs and to classify existing research in terms of evaluation appropriateness. Such a classification can be useful to new researchers and practitioners that aim to identify a scientific method appropriate for evaluating their PIAs and PRAs techniques. At the same time, industries and practitioners can identify and select PIAs and PRAs that have been rigorously evaluated based on the classification and comparison resulted from an SLR.

Research Questions

In order to study the evaluation and/or validation of PIAs and PRAs methodologies, we set to answer the following questions:

- **RQ1: What are the existing PIA and PRA techniques published in the scientific literature?**
    - The objective of this RQ is to identify available PIA and PRA techniques and classify them in terms of how they are operationalized, as well as in terms of their evaluation or validation.
- **RQ2: How and to what extent were these PIA and PRA techniques scientifically validated or evaluated?**
    - The objective of this RQ is to classify the identified PIA and PRA techniques by appraising their maturity in terms of whether they have passed through scientific rigorous evaluation strategies.

Methods

Design

This SLR follows the guidelines established by (Kitchenham, 2004). Nonetheless, for writing the SLR protocol, we also consider the preferred reporting items of protocols as proposed by (Moher *et al*, 2015) in addition to (Kitchenham, 2004).

Eligibility Criteria

Based on the research questions, and to reduce the likelihood of bias, the following inclusion and exclusion criteria are followed.

Inclusion	**Rationale**
Studies that propose a PIA/DPIA or PRA technique.	Identified publications need to contain a proposal of a new or improved technique for either PIA/DPIA or PRA.
Studies that described the validation or evaluation of a PIA/DPIA or PRA technique.	Identified publications need to describe or document a specific validation or evaluation strategy, e.g., expert interview that they use to validate their PIA/DPIA or PRA methodology.
Peer-reviewed studies/publications	This SLR is focused on the existing scientific peer-reviewed evidence on the topic. Studies that have not been properly peer-reviewed, e.g., book chapters, will not be included.

Exclusion
Papers not written in English
Studies/publications on PIA that do not analyze the method, specifically PRA
Studies/publications that enumerate or identify privacy risks

Studies/publications that focus on security analysis
Studies/publications do not critically analyze a risk methodology
Studies that are of low quality – no research question or clear methodology
Non-peer-reviewed studies
Studies/publications that do not perform apriori risk analysis

Information Sources and Search Strategy

For this SLR, we will search for publication in four scientific databases: Scopus, Web of Science, IEEE Xplore and ACM Digital Library. Google Scholar does not provide necessary elements for systematic scientific literature retrieval such as tools for incremental query optimization, export of a large number of references (Giustini *et al*, 2013), lacks Boolean search operations and the queries have been found to be irreproducible over time (Gusenbauer *et al*, 2020).To generate our main search string, the research question is broken down into individual facets. The table below shows that terms to be considered in the construction of a search string.

Table 1. Search terms organized in facets.

Individual facets	Synonyms	Key Terms
Privacy risk analysis and Privacy impact assessment	- Privacy risk assessment - Privacy risk analysis methodology - Privacy impact assessment - Data protection impact assessment	- Data protection risk assess* - Privacy risk assess* - Privacy risk analys* - Privacy impact assess* - Data protection impact assess* - Privacy threat model* - Privacy threat assess*

The generic search string is structured as follows:

Table 2. Generalized search string for the SLR.

Search string
("privacy impact assess" OR "privacy impact analys*" OR "privacy impact model*" OR "privacy risk assess*" OR "privacy risk analys*" OR "privacy risk model*" OR "privacy threat assess*" OR "privacy threat analys*" OR "privacy threat model*" OR "data protection impact assess*" OR "data protection impact analys*" OR "data protection impact model*" OR "data protection risk assess*" OR "data protection risk analys*" OR "data protection risk model*" OR "data protection threat assess*" OR "data protection threat analys*" OR "data protection threat model*")*

The search string will be further adapted for databases that have different search restrictions, for example IEEE, and if this will be the case, the search string will be indicated as well. The search strings used in each databases will be documented together with the total number of results retrieved and the dates of the searches. After the initial selection of studies, we will also follow the guidelines for snowballing as described by (Wohlin, 2014), i.e., backward and forward snowballing.

Study Records

Data management

To manage the screening process, we will export search results from each database and then import them to the RAYYAN software, allowing two reviewers to select papers independently (i.e., double-blinded) and manage conflicts by a third reviewer. Duplicated publications can also be removed using RAYYAN during the selection process. Bibliographies of final results will be exported to Zotero (for citing and sharing research).

Data extraction

Preliminary components to be extracted

- Bibliographic information, such as title, abstract, authors and affiliations, venues, year of publication, keywords, etc.
- Key information of the PIA or PRA techniques, such as its name, scope of analysis, type of risk analysis (qualitative or quantitative), association with a legal framework or standard.
- Validation and/or evaluation strategies used for the proposed PIA or PRA technique.
- Extent of evaluation – scale of validation activity that is measured e.g., number of survey, expert interviews
- Quality assessment and critical appraisal of the studies that have validated or evaluated a PIA or PRA technique.
- Target users of the PIA or PRA techniques, such as DPO, privacy practitioners etc., and the sector in which the PIA or PRA has been validated in.
- Information as to whether the PIA or PRA techniques assess privacy harms or how they conceptualize risks.

Data synthesis

Collating and summarizing results into classification tables and present the aggregated information in diagrams. We will also compose a narrative synthesis focused on the studies that have validated or evaluated their proposed PIA or PRA techniques.

Preliminary components of the data synthesis

- Overall identification and classification of PIA and PRA techniques in published research.
- Classification tables presenting validation and evaluation strategies applied for assessing PIA and PRA techniques.
- Comparison analysis – based on the magnitude of evaluation per PIA and PRA techniques.
- Narrative synthesis focused on the validated or evaluated PIA and PRA techniques.

Dissemination Strategy

Results from the research will made available through publications to conferences and in a PhD thesis. In addition, they will be disseminated through presentation to a group of network, for example, Hälso- och sjukvårdens informationssäkerhetsnätverk (HoSIS)

Conclusion

The systematic literature review will be used to identify the most relevant research for the research questions provided. It will provide a way to evaluate, collate and provide a

classification of available scientific methods used to evaluate PIA and PRA techniques. In addition, a comparison of the results can be provided to understand the magnitude of evaluation and provide an insight into which evaluation method is most likely preferable.

References

Kitchenham, B. (2004) 'Procedures for Performing Systematic Reviews'.

Vemou, K. and Karyda, M. (2018) 'An Evaluation Framework for Privacy Impact Assessment Methods', p. 11.

Wadhwa, K. and Rodrigues, R. (2013) 'Evaluating privacy impact assessments', *Innovation: The European Journal of Social Sciences*, 26. Available at: https://doi.org/10.1080/13511610.2013.761748.

Wohlin, C. (2014) 'Guidelines for snowballing in systematic literature studies and a replication in software engineering', in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. New York, NY, USA: Association for Computing Machinery (EASE '14), pp. 1–10. Available at: https://doi.org/10.1145/2601248.2601268.

Wright, D., Finn, R. and Rodrigues, R. (2013) 'A Comparative Analysis of Privacy Impact Assessment in Six Countries', *Journal of Contemporary European Research*, 9(1). Available at: https://doi.org/10.30950/jcer.v9i1.513.