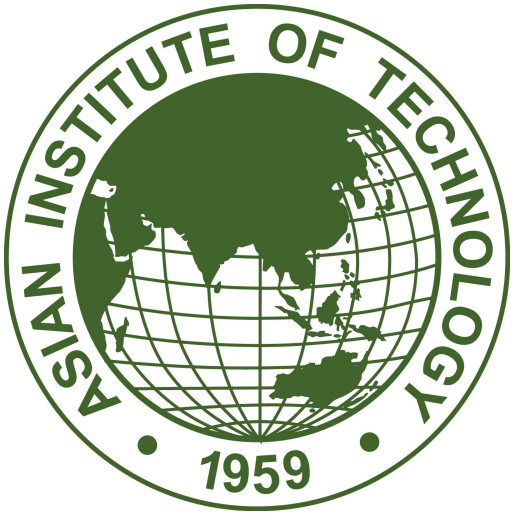
# Towards a privacy impact assessment methodology to support the requirements of the General Data Protection Regulation in a big data analytics context: A systematic literature review

by

**Sonu Adhikari, st124409**

A Summary of Paper Reading Presentation




**Asian Institute of Technology**

**School of Engineering and Technology**

**Thailand**

**April, 2024**

# Introduction

The article discusses the growing importance of big data analytics in today's business and organizational landscape. It highlights how big data analytics allows for the processing and utilization of large volumes of data, which can lead to improved efficiency and new opportunities. However, it also raises concerns regarding the increased risk of compromising or breaching personal data.

The General Data Protection Regulation (GDPR) is introduced as a regulatory framework that mandates organizations to conduct Data Protection Impact Assessments (DPIAs). These assessments aim to identify and mitigate risks associated with the protection of personal data. The article acknowledges that conducting DPIAs in the context of big data analytics is still relatively unexplored.

# Research Methodology

To address the gap of DPIAs in the context of big data analytics, the authors conducted a systematic literature review. They examined 159 articles and applied thematic analysis to identify privacy and data protection risks specific to big data analytics. This analysis resulted in the definition of nine Privacy Touch Points, which summarize the identified risks. The answer to the first research question resulted in the identification of 4 data privacy risks and 5 data protection risks which is mentioned in the points below.

**Table 1 – Research questions and their motivation.**

| ID | Research Question | Motivation |
|---|---|---|
| RQ1 | What are the specific privacy and data protection risks for Big Data Analytics? | We need to know the privacy and data protection risks specific to Big Data Analytics in order to assess the potential of PIA methodologies to be applied to systems using Big Data Analytics. |
| RQ1.1 | What are the privacy risks in the Big Data Analytics context? | We divided RQ1 in two sub-questions, acknowledging that privacy and data protection are not identical and have areas where their scope diverges, just as there are areas where their scope overlaps (see section 2 on the background for our study). |
| RQ1.2 | What are the data protection risks in the Big Data Analytics context? | For determining the privacy risks, we used as a guide the taxonomies of (Clarke, 2014) and (Finn et al., 2013). The latter because it adds two privacy types pertinent to recent technological advances. For data protection risks, we largely relied on Recital 75 of the GDPR. |
| RQ2 | To what extent do the PIA methodologies that have received attention in the reviewed literature, cover the risks identified in RQ1.1 and RQ1.2? | After having determined the referenced PIA methodologies in our publications sample, we identified how well these methodologies address the assessment of privacy risks and data protection risks specific for Big Data Analytics. |

Data Privacy Risks:

- Lack of transparency
- Treatment of indirect privacy harms
- Improper treatment of different types of privacy risks and data breaches
- Practical issues due to procedural vagueness

Data Protection Risks:

- Unclear data controllership
- Identification of individuals from derived data
- Discrimination issues affecting moral (e.g. stigmatization) or material (e.g. reducing the chance of finding a job) personal matters
- Increased scope leading to further processing incompatible with the initial purpose
- Limited range of stakeholders' involvement

The next step of the study involved analyzing ten existing Privacy Impact Assessment (PIA) methodologies which is addressing research question 2. The authors evaluated these methodologies to determine how well they cover the identified Privacy Touch Points. This analysis aims to provide insights into the strengths and weaknesses of current PIA methodologies in addressing privacy and data protection risks in the context of big data analytics.

**Table 5 – Extent of coverage of PTPs by PIA methodologies.**

| | More privacy risk related | | | | More data protection risk related | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | PTP (4) | PTP (6) | PTP (8) | PTP (9) | PTP (1) | PTP (2) | PTP (3) | PTP (5) | PTP (7) |
| CNIL | + | + | + | – | • | • | • | • | • |
| UK ICO | + | • | • | • | + | + | + | • | • |
| ISO/IEC 29,134:2017 | • | • | + | – | • | • | – | • | + |
| OPC | • | – | – | • | – | – | – | • | • |
| OAIC | • | + | • | • | • | • | • | • | • |
| US Homeland Security | – | – | • | – | • | • | – | • | • |
| NZ Privacy Commissioner | – | • | • | – | • | – | – | – | • |
| LINDDUN | + | + | • | – | – | + | + | + | • |
| IRL Data Protection Commission | • | – | – | – | • | • | • | • | • |
| HK PCPD | – | – | – | – | – | – | – | – | – |

(+) PTP documented or addressed in core process or supporting material.
(•) PTP referred to but not sufficiently addressed.
(-) PTP neither documented nor implied.

# Findings

From the above research analysis, we can see that three PIA methodologies stand out.

The PIA methodology of CNIL provides a good coverage of most privacy related PTPs, but not of the more data protection related PTPs. Exactly the opposite is true for the UK ICO PIA methodology. The LINDDUN privacy engineering approach provides the broadest coverage with two out of four privacy related PTPs and three out of five data protection related PTPs sufficiently covered.

The seven other methodologies assessed barely cover any specific privacy or data protection risks that are specific to the Big Data Analytics context. Overall, no single PIA methodology appears capable of addressing all Big Data Analytics-specific privacy and data protection risks.

# Conclusion

The insights gained from the analysis of the literature review and PIA methodologies will inform the subsequent phase of the authors' research. Their objective is to develop a comprehensive DPIA methodology that can assist data processors and data controllers in identifying, analyzing, and mitigating privacy and data protection risks specifically related to the storage and processing of data in big data analytics.

The article emphasizes the importance of complying with the GDPR's requirements, particularly the implementation of DPIAs, to ensure the appropriate protection of personal data and safeguard individuals' rights. By employing a robust DPIA methodology, organizations can proactively address privacy concerns and mitigate risks, thereby fostering a privacy-aware approach to big data analytics.